BLS WORKING PAPERS

U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Employment and Unemployment Statistics

Estimating the Civilian Noninstitutional Population for Small Areas
A Modified Cohort Component Approach Using Public Use Data

Andrew C. Forrester,
U.S. Bureau of Labor Statistics

Working Paper 571
September 2023

Estimating the Civilian Noninstitutional Population for Small Areas

A Modified Cohort Component Approach Using Public Use Data

Andrew C. Forrester*

U.S. Bureau of
Labor Statistics

September 2023

Abstract

This article develops a demographic method to estimate the civilian noninstitutional population for counties and county equivalents in the U.S. While these data provide the key sampling frame for national labor market surveys and denominators for labor market prevalence rates, the data are thus far unavailable for small areas. I develop a modified cohort component method to produce novel, monthly estimates of the civilian noninstitutional population for all U.S. counties using publicly available data on population and vital statistics with minimal modifications. The resulting population data may be used by researchers and policymakers to study within-year population dynamics as they relate to economic and demographic factors. I further extend the method to produce short-term population projections that include the most current vital statistics. The method compares favorably to existing annual, midyear estimates by the U.S. Census Bureau, but is prone to error in areas with fewer vital events.

JEL Codes: J21, J11, C81

Keywords: population estimation and projections; model specification; small areas

*Forrester: forrester.andrew@bls.gov, Division of Local Area Unemployment Statistics, Office of Employment and Unemployment Statistics, U.S. Bureau of Labor Statistics. I am grateful for comments from Sean Wilson, Tom Krolik, Luke Rogers, Larry Sink, and discussions at the Bureau of Labor Statistics and the Census Bureau. All views are my own and do not necessarily reflect those of the U.S. Bureau of Labor Statistics, the Department of Labor, or the United States.

1 Introduction

Labor force statistics in the United States rely upon estimates of the civilian noninstitutional population, which defines the sample frame for most socioeconomic surveys, such as the Current Population Survey (CPS). These data provide a key input to weight respondents to produce national and sub-national labor force statistics. While the U.S. Census Bureau produces intercensal and postcensal estimates of the civilian noninstitutional population at the national and state level, equivalent data are unavailable for smaller geographic areas, such as counties and county equivalents. This research proposes a demographic method to estimate the civilian noninstitutional population for counties and equivalents in the U.S. by combining publicly available data on population and vital statistics. The resulting data are available with monthly frequencies and permit researchers and planners to study local labor market dynamics with a high level of geographic and temporal granularity. Additional uses include their application to federal statistics production, as they match the population universe used to compute labor force measures for larger geographic areas. Finally, I apply the method to produce short term, monthly population projections.

To estimate labor force statistics and demographic characteristics of the population, the U.S. Bureau of Labor Statistics and the U.S. Census Bureau rely on the CPS, which is designed to reflect the civilian noninstitutional population ages 16 and older (CNP16) (U.S. Census Bureau, 2002). The U.S. Census Bureau produces estimates of the CNP16 to serve as independent population controls for the CPS and various other federal surveys (Land & Hough, 1986; U.S. Census Bureau, 2002). These population estimates are broken down by age, sex, race, ethnicity, and state of residence, providing the basis for labor force statistics, particularly as population denominators for labor force participation rates or employment to population ratios.

While these data are readily available for states and the nation as a whole, to date there are no comprehensive, publicly available estimates of the civilian noninstitutional population for smaller geographies, such as counties and county equivalents. While other estimates exist from the American Community Survey (ACS), they provide limited geographic coverage on an annual basis.[1] The only complete set of data is from the decennial census enumerations, which provide a complete count of the civilian resident population and the institutional population residing in group quarters facilities. Such institutional populations include inmates in prisons and jails, retirement and nursing homes, medical institutions, and hospices among others. These populations are crucial to measure, as they represent individuals who are unlikely to be attached to or participate in the labor force.

The objective of this article is to fill in this data gap by providing a unified method to produce monthly estimates of the civilian noninstitutional population for small areas. The method modifies the standard cohort component methodology by synthesizing monthly demographic components of change and institutional prevalence rates using publicly available data. Further calibrations link the population series over time before removing the institutional group

[1] The ACS provides limited information on the civilian noninstitutional population through Table S1811, which tabulates data on disability status. Coverage is limited to relatively few counties in both the 1-year and 5-year ACS data. I provide a comparison with the ACS CNP16 data in the Online Appendix.

quarters and military population to form local estimates of the civilian noninstitutional population. The resulting intercensal and postcensal data align to the population universe used to create standard labor force statistics and are especially important for researchers and planners looking to study labor force dynamics and population trends with both high frequency and at a more local level.

2 Background

The two key components of this research involve methods to produce population estimates for smaller time intervals and to adjust them to match the civilian noninstitutional definition. In each case, there is relatively limited literature focusing either on methods to produce higher frequency population estimates or how to estimate the civilian noninstitutional share of the population. While the U.S. Census Bureau produces these data by state and month to control the Current Population Survey (CPS), there are no existing methods applied to small areas (U.S. Census Bureau, 2002).

Developing intra-year population estimates for local areas is relatively sparse in the demography literature, particularly for small areas, such as U.S. counties and equivalents. Under ideal conditions, monthly population estimates would rely on well-measured components of demographic change arising from vital events (births and deaths) and in- and out-migration (Wilson et al., 2022). For small areas, estimating the resident population is often confounded by the lack of available data and requires analysts to impose assumptions on the demographic components of change to shift the population forward each period (Hauer, 2019). Producing monthly statistics of the population for small areas therefore carries a unique set of challenges regarding how the population changes each month, how to classify residents with demographic detail, and how to make best use of the data available.

Fundamentally, the demographic components of population change are subject to seasonal change resulting from seasonal residency, over-the-year variation in fertility and mortality, and migration patterns (Lam & Miron, 1996; Perz, 2004; Rosenberg, 1965). For example, many countries experience a spike in births during summer months due to an increase in conceptions during winter months (Lam & Miron, 1996). Similarly, seasonal patterns are frequently visible in death rates, as many countries show increases in death rates during winter months resulting from colder temperatures and higher levels of infectious disease (Feinstein, 2002). However, most estimates and projections, such as those produced by the U.S. Census Bureau, are made on an annual basis and reflect an area's midyear population.

Extant work on producing intra-year population estimates typically focuses on estimating temporary populations, which includes tourists, students, migrant workers, and business travelers. The size of the temporary population can vary greatly from one area to another based on local economic and geographic characteristics rather than the flow of population from vital events and migration. Since temporary populations are often not captured in censuses or surveys, Smith (1989) provides the first description of methods to develop estimates of temporary residents, comprising direct or indirect methods. Whereas direct methods require information collected from temporary residents themselves, indirect methods identify symptomatic variables that relate to an area's temporary population. Such approaches use administrative data such as residential electric customers (RECs), retail sales, sales tax receipts, and hotel occupancy rates to estimate temporary populations. Some of the earliest examples of the direct method utilized directly collected survey data from temporary residents. Looking at data on temporary residents in Florida through a consumer survey, Galvez & McLarty (1996) found seasonal peaks in Florida's population in winter months attributable to temporary residents. Comparable work in Arizona by McHugh et al. (1995) found similar seasonal peaks in winter months.

For the purpose of labor force estimation, temporary populations are less relevant compared to the resident population, which matches the residency requirement to tally employment and unemployment. Similarly, while temporary populations are likely to engage with an area's local economy, they are also less likely to be engaged in the local labor force (McHugh et al., 1995). Since local labor market analysis requires information on the population that *resides* in the area to align with the employment concepts used by the U.S. Bureau of Labor Statistics to describe household employment and unemployment (U.S. Bureau of Labor Statistics, 2022), methods designed to estimate the temporary population are less appropriate in this research. Instead, I turn to the method of cohort components to develop small area population estimates.

An important caveat to any use of the cohort component framework is the availability of quality data on vital events and migration for sub-national areas (Hauer, 2019; Wilson et al., 2022). In the case of U.S. counties and equivalents, it is often difficult, if not impossible, to obtain complete information about each demographic subprocess, especially with an appropriate level of demographic granularity (Hauer, 2019). With access to administrative data on vital events and migration, the Census Bureau fills in data gaps through its Administrative Record (ADREC) method to compute the demographic components of change for the balancing equation (Smith & Mandell, 1984; U.S. Census Bureau, 2009, 2021b). With these data constraints in mind, I develop methods to approximate each demographic component of change using available public data.

A further consideration is how to separate out the institutional group quarters and military populations from the resident population. Group quarters populations are especially difficult to measure, since the populations do not change according to the traditional demographic components of change and data are often available only through the decennial census. Consequently, the standard practice to estimate group quarters populations has been to carry the group quarters population from the most recent census forward, broken down by facility type and age unless more recent data are available (Bryan, 2004; U.S. Census Bureau, 2009, 2021b). Work by Land & Hough (1986) directly studied whether this approach yields valid estimates for the institutional group quarters population, finding substantial variation in the age distribution for institutional population over a long time frame (1940-1980). Land & Hough (1986) leverage detailed data on the age distribution of the civilian institutional population to compute institutional prevalence rates for the population. The authors further suggest that such rates may be applied to intercensal and postcensal estimates to improve their accuracy when estimating the institutional population, and therefore, the civilian non-institutional population.

Similar methods to separate out the institutional and military populations also exist; however, they are often based on administrative data specific to particular state laws. For example, the State of Alaska utilizes administrative records derived from enrollment in the Alaska Permanent Fund Dividend (PFD) to track population and migration (Alaska Department of Labor, 2021). Comparatively, the State of California uses administrative records from its driver's license database and Medi-Cal health insurance enrollment to produce its population estimates (State of California, Department of Finance, 2023). Comparatively, these methods are untenable to implement on a national basis, as they blend numerous disparate methods, require detailed institutional knowledge for each state, and are predicated on access to the respective administrative databases.

Ultimately, my research is closely related to work by Hauer (2019), which develops a uniform method to produce long-term population projections for U.S. counties and equivalents and controls them to independent projections for the nation. Hauer (2019) uses the Hamilton & Perry (1962) method to produce long-term population projections, with a high degree of demographic detail, including age group, sex, race, and ethnicity. The Hamilton-Perry method provides a reliable, data-driven framework to project population by applying cohort change ratios (or level differences) to a base population and bypasses the need for vital statistics and migration estimates. While the Hamilton-Perry method has minimal data input requirements, it does not incorporate demographic components into its projections and therefore is less apt to model the seasonal variability in the overall population. Comparatively, my objective is to provide higher frequency population estimates and projections for areas with a relatively less demographic detail, focused on the civilian noninstitutional population, specifically.

3 Data

3.1 Population Data

The U.S. Census Bureau produces midyear population estimates by age, sex, race, and ethnicity and their respective components of change for counties and county equivalents. These data are produced by the *Population Estimates Program (PEP)* and use administrative data collected by the U.S. Census Bureau and through the Federal and State Cooperative for Population Projections (FSCPE) (U.S. Census Bureau, 2021b). Administrative records provide the main data inputs into a standard cohort component model to produce population estimates each year based on the most recent census enumeration, known as the administrative record, or ADREC, method (Smith & Mandell, 1984).

I rely primarily on two datasets from PEP. The first dataset is the AGESEX database, which provides estimates of the resident population by age group and sex for states, counties, and equivalents.[2] From these data, I take the total population age 15 and ages 16 and over (henceforth referred to as "ages 16 plus") by county and equivalent and year. The second database is the Components of Change files, which contain the demographic components of

[2] These data are available from the U.S. Census Bureau's FTP site (U.S. Census Bureau, 2022a, 2022b).

change for the total population, the rates of population change, and the total group quarters population.[3] Using these data, I collect the total net migration rates and total group quarters estimates for each county and equivalent by year.

While each dataset fundamentally provides the same information, the population estimates data come from two separate releases, or vintages, for 2020 and 2021. The Vintage 2020[4] data span April 2010 through July 2020 and use the 2010 enumeration as a population base, while the Vintage 2021 cover April 2020 through July 2021 and rely on a "blended base" instead. Resulting from delays and issues in the 2020 enumeration, the Census Bureau developed a blended population base for its subsequent population estimates that synthesizes data from the 2020 enumeration, the 2020 Demographic Analysis, and the Vintage 2020 population estimates. Consequently, there are observable differences between the postcensal estimates from Vintage 2020 and Vintage 2021 with a reference date of April 2020. Until the Census Bureau releases an official intercensal series for 2010 through 2020, I use the Das Gupta (1981) method to link the two vintages (U.S. Bureau of Economic Analysis, 2022).

3.2 Vital Statistics

Vital statistics from the National Center for Health Statistics (NCHS) consist of monthly births and deaths by county and equivalent and are compiled through the National Vital Statistics System. Each dataset is subject to release constraints set by the NCHS to prevent data disclosure.

Data on births represent a register of all birth certificates in the 50 states and the District of Columbia (Osterman et al., 2023). For data privacy concerns, births are suppressed for areas with a population under 100,000. For suppressed areas, the NCHS provides a "Balance of State" count of births per month. The population threshold for data release floats based on population counts from the most recent census enumeration. Consequently, the component areas of each balance of state change discretely when the NCHS revises its population thresholds. To avoid bias arising from compositional effects when estimating the birthday shares, I normalize all births data into their base year groupings, i.e., recursively combine all newly enumerated areas into balance of state estimates back in time. I source the monthly births data by tabulating the 1994-2006 microdata by area and birth month.

Mortality statistics tabulate death certificate records for all 50 states and the District of Columbia (Xu et al., 2022). These vital statistics contain the number of deaths by county and equivalent, year, month, and the decedent's age. Using the CDC WONDER system, I subset the data to include deaths for individuals age 16 plus spanning 2010 to the most recent month. While the NCHS data are administrative records on deaths, there are important caveats to the NCHS data that are worth noting.

First, the deaths data are subject to release limitations to protect individuals' privacy. The NCHS suppresses all monthly death counts where the monthly death count is below 10 deaths

[3] These data are available from the U.S. Census Bureau's FTP site (U.S. Census Bureau, 2021a, 2022c).

[4] These data are also called the evaluation estimates and span April 2010 through July 2020. The evaluation estimates are postcensal estimates based on the 2010 enumeration and do not incorporate information from the 2020 enumeration.

(Hauer, 2019). Consequently, not all area-months contain deaths data. When the data are censored, I provide a simple imputation to fill in missing data by allocating the annual total to each missing month according to how many months are missing.[5] For most states, the remainder is very small and the counties and equivalents that have missing deaths data are few in number. Further research might examine more sophisticated models to estimate the suppressed death counts, although the impact of data suppression on the population estimates is small given the low supression threshold.

Second, the death counts for the two most recent years are provided as provisional counts. With each year, the NCHS corrects and updates the provisional data to produce final estimates on a two-year lag. Despite their provisional nature, the U.S. Census Bureau uses the provisional deaths for its population estimates in its administrative record approach (U.S. Census Bureau, 2021b). Following the U.S. Census Bureau, I implement a simple forecasting method for each county and equivalent's mortality time series to implement short-term projections (Hauer, 2019; U.S. Census Bureau, 2021b).

Third, the mortality statistics are likely impacted by excess mortality brought on by the COVID-19 Pandemic. Comparing expected mortality for counties based on pre-pandemic trends, Paglino et al. (2023) found that cumulative excess mortality brought on by the Pandemic was concentrated in nonmetropolitan counties where mortality data are often less complete due to data suppression. Since data suppression constraints also exist for COVID-19-related deaths, it is difficult to quantify the extent to which distortions from COVID-19 will impact the subsequent population estimates. Further discussion will discuss how excess mortality is incorporated in the estimates.

3.3 Group Quarters

The group quarters (GQ) population includes persons residing in common living quarters, such as prisons, nursing homes, college dormitories, and military barracks, among others. The most complete data available for the GQ population by broad age group and facility type come from the decennial enumeration. Specifically, I use data from Summary File 1 from the 2010 Decennial Census and the Demographic and Housing Characteristics (DHC) file from the 2020 Decennial Census for all counties and equivalents in the U.S.[6] The GQ population is available in five-year age ranges for seven GQ facility types. I collect GQ population by age and sex for the total GQ population (PCO1), institutionalized population (PCO2), and military quarters (PCO9). The 2010 census did not collect information on civilian or armed forces status; rather, the 2010 enumeration tabulates resident military within military quarters (U.S. Census Bureau, 2012a). I therefore use the population residing in military quarters as a proxy for the resident military population by county and equivalent.

[5] The NCHS data-use restrictions forbid publication of counts or death rates based on fewer than 9 deaths.

[6] The American Community Survey (ACS) 5-year estimates also provide the same information; however, the 5-year estimates also lack timeliness and require additional assumptions when comparing across ACS 5-year samples. Additionally, the ACS tabulates the data by age using unconventional age groups that would require more assumptions to arrive at estimates of the population ages 16 plus.

4 A Modified Cohort Component Method

In this section I describe a demographic method to compute monthly estimates of three populations for counties and equivalents. The process involves estimating three populations: (i) the resident population ages 16 plus, (ii) the institutional and resident military population ages 16 plus, and (iii) the civilian noninstitutional population ages 16 plus. I first estimate the monthly resident population by applying synthesized components of change to the base population in each estimation period. I then introduce a prevalence rate method to compute the institutional and resident military population ages 16 plus each month. Finally, I compute the civilian noninstitutional population ages 16 plus by subtracting the institutional and resident military population from the resident population. To reconcile the county-level estimates with official estimates from the U.S. Census Bureau, I use two methods to adjust and control my estimates to official data sources. In the subsequent discussion, I refer to "estimates" and "projections" interchangeably, as I rely on a combination of historical and synthetic data to produce the monthly population series.

4.1 Estimating the Resident Population

First, I compute the resident population ages 16 plus for each month by applying the population balancing equation (Bryan, 2004; Preston et al., 2000; U.S. Census Bureau, 2002). The balancing equation produces estimates, or projections, of the resident population as the cumulative sum of the demographic components of change applied to a base population. The components of change characterize the increase or decrease of the population and include adding births, subtracting deaths, and accounting for in- and out-migration (Preston et al., 2000). Specifying the balancing equation ultimately depends on the desired population universe and requires disaggregating each component of change accordingly. Since the target population for this research is the resident population age 16 and over, I adjust the standard cohort component framework to account for monthly aging from 15 to 16, deaths to those 16 plus, and net migration for those 16 plus.

For the representative area i observed over two periods, denoted t and $t+1$, the resident population follows the general balancing equation

$$\texttt{POP16}_{i,t+1} = \texttt{POP16}_{i,t} + \texttt{AGING16}_{i,t} - \texttt{DEATH16}_{i,t} + \texttt{NETMIG16}_{i,t}, \tag{1}$$

where:

- $\texttt{POP16}_{i,t}$ is the resident population ages 16 plus
- $\texttt{AGING16}_{i,t}$ is the population aging into the 16 plus age group over the month
- $\texttt{DEATH16}_{i,t}$ is the number of deaths for the population ages 16 plus
- $\texttt{NETMIG16}_{i,t}$ is the level of net migration for ages 16 plus

- t indexes months in order between the base month and the target month such that $t \in \{Jul, Aug, \ldots, May, Jun\}$

To compute the population each month, I apply the balancing equation iteratively for each month throughout the estimation period. I define the estimation period as the July 1st estimate in the first year (the base month) and July 1st of the subsequent year (the target month). Starting with the base month, I move the population age 15 ahead into the 16 plus group by applying the share expected to turn 16 each month. From this preliminary 16 plus population, I subtract out the number of deaths for those ages 16 and over to account for the natural increase in the population. Finally, I compute net migration by taking the aged population as the population at-risk to migrate and apply a monthly net migration rate to determine the level of net migration for the population ages 16 plus. I then iterate this process each month until I reach a population estimate for July 1st in the subsequent year.

The iterative approach requires estimates for each of the three demographic components of change over the course of the year. The following discussion describes each of these processes in detail.

4.2 Measuring the Components of Change

Estimating the balancing equation in Equation 1 for each estimation period requires a base population and the demographic components of change that match the resident population ages 16 plus universe. Further, the proposed method requires monthly estimates of the components of change. The following discussions describe methods to estimate each component at a monthly frequency that matches the resident population 16 plus universe.

4.2.1 Aging Process

Following the U.S. Census Bureau (2021b) methodology, I take the July 1st (or April 1st in decennial years) population age 15 and move a portion of them forward one year of age each month. This process requires an estimate of how many 15-year-olds turn 16 in each month of the year. To synthesize this component on a monthly basis, I use data on the historical distribution of births by month sourced from the NCHS. I measure monthly aging by adding a portion of 15-year-olds in the base month into the 16 plus population relative to the historical share of births in the area each month. Combining the two components provides a monthly time series for the number of 15-year-olds turning 16 each month of the year.

For each area i, I compute the expected share of birthdays in each month using the historical distribution of births $B_{i,m,y}$ by year y and month m. Given a known total of births in an area each year, I approximate the share of births in each month as the probability $\delta_{i,m,y}$ of being born in month m in year y following

$$\delta_{i,m,y} = \frac{B_{i,m,y-16}}{\sum_m B_{i,m,y-16}} \qquad \forall\ y = 2010, \ldots, T. \tag{2}$$

The set of estimated $\widehat{\delta_{i,m,y}}$ represent the share of the birthdays over the course of month m. Using the total number of 15-year-olds from the base month, I apply the shares to the proportion of the population age 15 in the base month to compute monthly aging into the 16 plus age group as follows

$$\texttt{AGING16}_{i,t} = \hat{\delta}_{i,m,t} \times \texttt{POP15}_{i,t_0}. \tag{3}$$

$\texttt{AGING16}_{i,t}$ is the population turning age 16 in month t moving into month $t+1$ and $\widehat{\delta_{i,m,t}}$ is the historical share of births that occurred in month t. The resulting aging series approximates the monthly number of 15-year-olds turning 16 and aging into the target population.

This process assumes that the share of people aging into the population ages 16 plus $\widehat{\delta_{i,m,t}}$ is well approximated by the share of people born in each month within the county 16 years ago.[7] This assumption may be violated for areas where the number of 15-year-olds changes substantially over time due to migration. For example, if a large number of teenagers move out of an area between the month they were born and the current estimation month, the historical distribution of birthdays would be less reflective of the current birthdays of 15-year-olds.

4.2.2 Mortality Process

The mortality process is relatively straightforward, as data on mortality are widely available from the NCHS subject to some data suppression. The NCHS publishes tabular records of death certificates by county and month, broken down by age and other characteristics. For my purposes, I use deaths for individuals ages 16 and older. These data come from two primary datasets: the final and provisional data. Since the final data are released at a lag, the NCHS also published provisional death counts with the same characteristics as the final data to bridge the gap from the final data release to present.

In the event that population projections are needed beyond the latest published month in the provisional data, I propose a similar method described in Hauer (2019) and use a simple univariate time series model to project deaths. For each area mortality time series, I fit an `ARIMA`(0,1,1) model to predict the remaining monthly death counts for the production year. The `ARIMA`(0,1,1) simplifies to simple exponential smoothing and is a parsimonious model to predict death counts. Since the latest month's death count is a provisional and incomplete count, I drop the latest partial observation for each area time series when fitting each `ARIMA` model. The resulting process creates a time series of monthly deaths for ages 16 plus by area through the projection target month. Further research would benefit from examining whether more complex time series models may improve short-term forecast mortality.

[7] An alternate data source on the intra-year distribution of birthdays is the American Community Survey (ACS) microdata, which records each respondent's quarter of birth. While the ACS provides a more current measure than historical births data, the geographic coverage is loosely comparable to that of the NCHS births data. The lowest identifiable geography in the microdata is the Public Use Microdata Area (PUMA) and covers geographic areas with a minimum population of 100,000 residents. Further research may examine whether the quarter of birth estimates from the ACS provide more reliable population estimates than the historical births data.

4.2.3 Net Migration Process

I adopt a rate-based approach to estimate the net migration component. This process relies on net migration rates from the U.S. Census Bureau's *Population Estimates Program* to compute net migration rates for the 16 plus population by month. The U.S. Census Bureau (2021b) uses internal tabulations of tax return data from the Internal Revenue Service (IRS), Medicare enrollment data from the Centers for Medicare and Medicaid Services (CMS), and the American Community Survey (ACS) to estimate net migration rates for each area. Migration rates for the population ages 0-64 rely on matched tax return data between two tax filing years, while the migration rates for ages 65 plus rely on Medicare administrative records (U.S. Census Bureau, 2021b).

To estimate monthly net migration for the population ages 16 plus, I use the net migration rates per 1,000 published in the PEP components of change data to develop a synthetic migration component similar to Wilson (2022). Specifically, I estimate the net migration component by allocating the annual net migration rate $\texttt{NMR}_{i,t_0}$ to each month. I then apply this monthly rate to the base month's aged population as follows

$$\texttt{NETMIG16}_{i,t} = \left(\texttt{POP16}_{i,t} + \texttt{AGING16}_{i,t}\right) \times \left(\frac{\texttt{NMR}_{i,t_0}}{12}\right), \tag{4}$$

where:

- $\texttt{NETMIG16}_{i,t}$ is the level of monthly net migration that occurs over t to $t+1$
- $\texttt{POP16}_{i,t_0}$ is the base population ages 16 plus
- $\texttt{NMR}_{i,t_0}$ is the annual net migration rate

This synthetic approach process makes two key assumptions. First, I assume that the net migration rate for the area's total population is the same for the population ages 16 and over.[8] Second, I assume that the net migration rate is stable across the entire year. While the second assumption may be less appropriate for areas with highly seasonal populations, such as college and vacation towns, there are no external data that provide within-year variation in area-to-area migration. Typical symptomatic variables used to measure seasonal populations, for example residential electric customers, are not universally available for all counties and therefore represents a data limitation. Similarly, since migration rates are determined by annual changes in either tax records or Medicare enrollment for domestic migration or residence one year ago for immigration, a higher frequency estimate for the migration process is not viable.

[8] The Population Estimates Program produces separate migration rates for two age groups, under 65 and 65 plus. The distinction between these two groups is the data source used to compute the migration rates. For the population under 65, the Census Bureau uses data on IRS tax filings; whereas, the 65 plus migration rates come from Medicare enrollment records from the Centers for Medicare and Medicaid Services (CMS). Since the CMS data are non-public, I use the total net migration rates produced by PEP for the entire population ages 15 plus.

The Census Bureau's PEP components of change also represent the best available data for migration compared to similar public use data from the IRS. Recent work by DeWaard et al. (2022) finds a major data quality issue with the IRS-based measures of county-to-county migration. The authors document systemic differences in migration rates post-2011 after the IRS took over the data production process from the U.S. Census Bureau, casting doubt on the data quality after the 2011 tax year. For this reason, I opt to use the migration rates directly published by PEP instead of the public data from the IRS Statistics of Income (SOI). Using additional information from the Social Security Administration (SSA) and the Centers for Medicare and Medicaid Studies (CMS), the Census Bureau develops more detailed migration estimates by internally matching IRS tax return data and SSA data (U.S. Census Bureau, 2021b). Further, the Census Bureau migration rates capture international in- and out-migration using data from the ACS residence one year ago (ROYA) question for immigration and survival rates for out-migration.

4.3 Group Quarters Population

Since the objective is to estimate the civilian noninstitutional population, the next step is to estimate the institutional and military group quarters populations to subtract from the estimated resident population each month (Land & Hough, 1986).

The first step is to estimate the age distribution for the institutional group quarters population (U.S. Census Bureau, 2021b). This involves computing the *institutional prevalence rate* for each age group, or the share of each age group residing in institutional or military quarters relative to the total group quarters population (Land & Hough, 1986; U.S. Census Bureau, 2009, 2021b). Since the data are available only in five year age ranges, I use the Beers (1945) 6-parameter method to interpolate the population for ages 16 through 19 to compute the population ages 16 plus residing in each facility type. The Beers (1945) method is a standard demographic method to create a smoothed single year age distribution from data reported in five year age groups.[9]

Once I compute the group quarters population 16 to 19 for each facility type, I add in the corresponding population ages 20 plus to arrive at the group quarters population 16 plus by facility type. Using these data, I estimate the *institutional prevalence rate* as the share of the group quarters population residing in institutional or military group quarters for each area i in census year y using

$$\texttt{INSRATE}_i^y = \frac{\texttt{GQINST16}_i^y + \texttt{GQMIL16}_i^y}{\texttt{GQTOTAL}_i^y} \qquad \text{for } y \in \{2010, 2020\}, \tag{5}$$

where:

- $\texttt{INSRATE}_i^y$ is the estimated institutional prevalence rate

[9] For areas with extremely small GQ populations, the Beers (1945) formula occasionally returns negative counts for the single age 16 population. In these cases, I bottom code the GQ population for age 16 as zero. Since the populations in question are so small, this edit has no substantive effect on the subsequent prevalence rate calculations.

- $\texttt{GQINST16}_i^y$ is the GQ population ages 16 plus living in institutional GQs
- $\texttt{GQMIL16}_i^y$ is the GQ population ages 16 plus living in military GQs
- $\texttt{GQTOTAL}_i^y$ is the total group quarters population

The resulting *institutional prevalence rates* for each area provide a means to adjust the annual total group quarters population estimates from PEP to arrive at the estimated institutional group quarters population ages 16 plus by area. I compute a set of rates using the 2010 and 2020 decennial censuses and linearly interpolate the prevalence rates between the two censal years to prevent a discontinuous jump in the prevalence rates when changing from the 2010-based rates to the newer 2020-based ones. Following the 2020 enumeration I carry the 2020-based prevalence rates forward for each month, assuming the most recent $\texttt{INSRATE}_i^y$ does not change going forward (Bryan, 2004; U.S. Census Bureau, 2021b).

To compute the institutional and resident military populations, I apply the institutional prevalence rate to the total group quarters population to arrive at the institutional and resident military population ages 16 plus.

4.4 Forming Consistent Time Series

After applying the synthetic components of change to the base population over each month, the final step is to create a linked resident population time series by adjusting the monthly cohort component estimates to reflect the error of closure between the cohort component method and the published PEP estimates. Linking the monthly estimates to the next year's July 1st (or April) estimates ensures that the resulting series are consistent across the entire intercensal population series and, therefore, allows comparisons within each area over time. Following the U.S. Census Bureau (2012b), I link each series for each year using the Das Gupta (1981) 6 factor method. This process allocates the difference between the modified cohort component estimate and the PEP estimate across each month over the course of the year. I then use the same method to link the PEP data across Vintage 2020 and 2021.

Taking the difference between the modified cohort component and official estimates, or the error of closure, I adjust each month over the preceding year using the Das Gupta (1981) method, the preferred method by the U.S. Census Bureau (U.S. Census Bureau, 2012b) and the Bureau of Economic Analysis (U.S. Bureau of Economic Analysis, 2022). The Das Gupta (1981) method assumes that the ratio of the modified cohort component estimates to the PEP official data progresses geometrically over the year. Letting T denote the terminal period and t_0 denote the first period, the general Das Gupta (1981) framework takes the form

$$P_{i,t} = Q_{i,t}\left(\frac{P_{i,T}}{Q_{i,T}}\right)^{\frac{t-t_0}{T}}, \tag{6}$$

where:

- $P_{i,t}$ is the intercensal estimate
- $Q_{i,t}$ is the postcensal estimate
- t_0 is the base period
- T is the terminal period

The ratio $P_{i,T}/Q_{i,T}$ represents the error of closure between the modified cohort component and PEP midyear estimates to distribute across months of the year. After adjustment, the resulting series will match the official PEP July 1st estimates while providing monthly variability in the population series.

A final consideration is the difference in area population observed while switching between PEP vintages 2020 and 2021. PEP releases "vintage" estimates that correspond to the data release year and the population base used to develop the midyear estimates. Owing to issues with data collection for the 2020 census enumeration, the Census Bureau opted to use a "blended base" method in their Vintage 2021 population estimates (U.S. Census Bureau, 2021b). The blended base approach was designed to overcome deficiencies in the 2020 enumeration by incorporating additional data from other data sources, including the 2010 census, the vintage 2020 population estimates, and other administrative records. Consequently, comparing the Vintage 2020 and 2021 estimates is inappropriate, since the population estimates are computed using different bases.[10]

In the absence of a linked intercensal series for 2010-2020, I adjust the Vintage 2020 postcensal series using the Das Gupta (1981) method.[11] In this approach, I distribute the difference between the Vintage 2020 April 1st postcensal estimate and the Vintage 2021 April 1st estimates base across the 2010-2020 time series. This final adjustment ensures that the Vintage 2020 based monthly estimates are directly comparable to the subsequent Vintage 2021 estimates and projections.

4.5 Civilian Noninstitutional Population

The civilian noninstitutional population is defined as the resident population less the resident armed forces and the institutionalized population (U.S. Census Bureau, 2002). With estimates of the resident population ages 16 plus $\texttt{POP16}_{i,t}$ and the institutional and military GQ population $\texttt{GQINS16}_{i,t}$, I produce the final estimates for the civilian noninstitutional population ages 16 plus $\texttt{CNI16}_{i,t}$. This process subtracts the institutional and military group quarters population from the resident population ages 16 plus as follows

$$\begin{aligned} \texttt{CNP16}_{i,t} &= \texttt{POP16}_{i,t} - (\texttt{INSRATE}_\texttt{i} \times \texttt{GQESTIMATE}_{i,t}) \\ &= \texttt{POP16}_{i,t} - \texttt{GQINS16}_{i,t}, \end{aligned} \tag{7}$$

where:

[10] The Census Bureau plans to release intercensal series for 2010-2020 in 2023.

[11] The same approach is used by U.S. Bureau of Economic Analysis (2022) to adjust the population data to prepare per capita personal income time series.

- $\texttt{CNP16}_{i,t}$ is the estimated civilian noninstitutional population ages 16 plus
- $\texttt{POP16}_{i,t}$ is the total population ages 16 plus
- $\texttt{INSRATE}_{i,t}$ is the institutional prevalence rate for ages 16 plus
- $\texttt{GQESTIMATE}_{i,t}$ is the total group quarters population

The resulting series is a an uncontrolled monthly estimate of the civilian noninstitutional population ages 16 plus by county and equivalent and month.

4.6 Controlling the Estimates and Projections

To ensure the CNP16 estimates align with the official data produced by the U.S. Census Bureau, I follow the official PEP methodology and other literature and rake the monthly estimates to a control CNP16 series (Hauer, 2019; U.S. Census Bureau, 2021b). Specifically, I control each county CNP16 series to its respective monthly state CNP16 series produced by the U.S. Census Bureau for the Current Population Survey (CPS) and published by the Local Area Unemployment Statistics (LAUS) program at the BLS (U.S. Census Bureau, 2002).

For each area i in state s and month t, I estimate the rake using the state control total as follows

$$\texttt{RAKE}_{i,t} = \left(\frac{\texttt{CNP16}_{s,t}}{\sum_{i \in s} \texttt{CNP16}_{i,t}}\right) \qquad \forall\ i, t. \tag{8}$$

With a rake factor computed for each area and month, I multiply the CNP16 value by the rake factor to compute the final controlled CNP16 estimate for each area. With only one control total for each state and month, the top-down controlling process is straightforward and ensures that the monthly area CNP16 estimates are additively consistent with the states and, therefore, the nation. Since the U.S. Census Bureau has not released an intercensal CNP16 series for 2010-2020, I use the Das Gupta (1981) method to reconcile the Vintage 2020 estimates with the Vintage 2021 blended base estimates and can be updated as new vintages are released.

In addition to addivity, there are two additional benefits to the control step. First, controlling to the independent population series minimizes measurement error in the institutionalized and military GQ populations, since the state-level estimates from the U.S. Census Bureau incorporate more current information about the group quarters and resident armed forces population (U.S. Census Bureau, 2002). Second, the independent, state-level population series prevent runaway population change resulting from measurement error in the vital statistics, for example Pandemic-related excess mortality for localities (Gonzalez-Leonardo & Spijker, 2022).

4.7 Short Term Projections

The modified cohort component framework provides a flexible means to develop short term population projections. I produce short term projections for the three population groups: (i) the resident population ages 16 plus, (ii) the institutionalized and resident military population, and (iii) the civilian noninstitutional population ages 16 plus.

Following the terminology in Rayer (2008), I begin the population projections at the "launch month," or the latest official midyear population estimate. I then use data on population aging, provisional and forecast mortality, and net migration rates to forecast areas' populations ahead iteratively until the target month. For the purposes of this study, I use the latest Vintage 2021 estimates ending in July 1st, 2021 (launch) to project the monthly area population through March 1st, 2023 (target).

To project the resident population ages 16 plus, I rely on provisional data on mortality, lagged historical data on the distribution of birth months, and historical trends in net migration rates. Using the balancing equation in Equation 1 and replacing each component with its estimate, I project the resident population using

$$\widehat{\texttt{POP16}}_{i,t+1} = \texttt{POP16}_{i,t} + \widehat{\texttt{AGING16}}_{i,t} - \widehat{\texttt{DEATH16}}_{i,t} + \widehat{\texttt{NETMIG16}}_{i,t} \qquad \text{for } t > T. \tag{9}$$

Similarly, I apply the balancing equation iteratively each month from the launch month through the target month. The resulting (uncontrolled) population series reflects the monthly resident population implied by projected aging, deaths, and net migration. This procedure incorporates similar data on vital statistics from the NCHS in the form of the intra-year distribution of birth months and provisional deaths for ages 16 plus. To compute the net migration component, I carry forward the net migration rates from the most recent population estimates vintage.

Computing the institutionalized and resident military population involves a simpler process. Since the demographic structure of the GQ population is often constant over time, I impose the same assumptions as the U.S. Census Bureau and carry the population from the launch month forward until the target month (Bryan, 2004; U.S. Census Bureau, 2021b). The institutionalized and resident military population therefore follows

$$\widehat{\texttt{GQINS16}}_{i,t+1} = \widehat{\texttt{GQINS16}}_{i,t} \qquad \text{for } t > T. \tag{10}$$

With projections for the resident and institutional and resident military populations, I compute the final projection for the civilian noninstitutional population as

$$\widehat{\texttt{CNP16}}_{i,t+1} = \widehat{\texttt{POP16}}_{i,t+1} - \widehat{\texttt{GQINS16}}_{i,t+1} \qquad \text{for } t > T. \tag{11}$$

In a final step, I rake the uncontrolled population projections to an independent monthly state CNP16 series from the CPS.[12] This raking procedure for the projections has two pri-

[12] An alternative control series is the monthly national CNP16 projections from the U.S. Census Bureau. I opt to use the projected CNP16 estimates from the CPS, as they represent the most current data available for CNP16.

mary benefits. First, controlling the county projections to the state level ensures that the county data will align with the CNP16 estimates used to compute the official, state-level labor force statistics. The population denominators for counties will therefore match the population universe of their state equivalents. Second, raking the projections to the state control series prevents runaway population growth for the county level projections. The final CNP16 series are consistent with the official statewide population denominators used by the BLS both historically and for the recent projections.

4.8 Error Measures

I use two primary measures to evaluate how well the modified cohort component method aligns with existing intercensal estimates. This entails measuring the accuracy and bias of the modified cohort component model compared to a "true population." While the literature typically relies on a subsequent census to evaluate estimation quality, I benchmark the monthly population estimates against the Census Bureau's midyear population estimates, i.e., the "true" population value. I rely on two commonly used accuracy measures, the Mean Absolute Percent Error (MAPE) and the Mean Algebraic Percent Error (MALPE) (Bryan, 2004; Smith & Sincich, 1990). For the following discussions, I refer to the official data from PEP in the target month T as $\texttt{POP16}_{i,T}$ and the modified cohort component estimate as $\widehat{\texttt{POP16}}_{i,T}$.

The MAPE measures the average deviation of the modified cohort component model from the official Census Bureau midyear estimates, regardless of the direction of the difference. A higher MAPE shows that the modified cohort component estimates tend to diverge from the official PEP estimates. I compute the MAPE as

$$\texttt{MAPE}_i = 100 \times \frac{1}{T} \sum_t \left| \frac{\texttt{POP16}_{i,T} - \widehat{\texttt{POP16}}_{i,T}}{\widehat{\texttt{POP16}}_{i,T}} \right| \qquad \text{for } i. \tag{12}$$

A related error measure that accounts for the direction of model error is the Mean Algebraic Percent Error (MALPE). The MALPE measures the extent to which the modified cohort component method over- or under-shoots the official PEP estimates. I compute the MALPE as follows

$$\texttt{MALPE}_i = 100 \times \frac{1}{T} \sum_t \frac{\texttt{POP16}_{i,T} - \widehat{\texttt{POP16}}_{i,T}}{\widehat{\texttt{POP16}}_{i,T}} \qquad \forall\, i. \tag{13}$$

Together, the MAPE and MALPE provide standard measures of both the magnitude and direction of the difference between the modified cohort component approach and the official U.S. Census Bureau population estimates for the population ages 16 plus. For each set of comparisons, I use the difference between the estimated and actual July 1st population estimates.

5 Results

In this section I present the population data and projections generated using the proposed cohort component method and provide various technical validations for the method. First, I show the resulting time series data developed through the modified cohort component method, both for the resident and civilian noninstitutional populations ages 16 plus. Then, I evaluate how well the modified cohort component method performs in two ways: its ability to match the estimates developed by the Census Bureau's ADREC method and how realistic the short term projections are. The first test checks whether the synthetic components of change and their underlying assumptions provide a good approximation of those used by the Census Bureau to develop their official estimates. For this purpose I draw on standard measures of fit to compare the modified cohort component method estimates with the official data from U.S. Census Bureau. The second test examines whether the proposed method provides reasonable out-of-sample population projections, comparing the forward projections against other standard projection methods.

5.1 Data Presentation

First, I visualize each step in the proposed methodology in Figure 1 for Autauga County in Alabama. These steps broadly include applying the synthetic components of change to the base population each year, controlling the final estimate to the official Census Bureau data, bridging the Vintage 2020 and 2021 data to form a synthetic intercensal series, and computing the final civilian noninstitutional population ages 16 plus (CNP16). Each panel highlights the population time series resulting from each step.

Panel A shows the first step of the modified cohort component method, in which I apply the synthetic components of change to the July 1st (or April 1st) population base in each year. The resulting "Modified CCM" time series, shown in blue, represents the uncontrolled estimates of the resident population ages 16 plus for each month. For comparison purposes, I highlight the official POP16 estimates from the Census Bureau with blue dots. Visually, the modified cohort component method looks to perform well in approximating the official population estimates each year, indicated by small differences between the modified cohort component series and the official estimates.

The next step in Panel B ensures that the monthly series aligns with the official Census Bureau estimate in the subsequent year. To reconcile the estimates with the official data, I control the monthly series to the next midyear estimate using the Das Gupta (1981) formula. This process allocates the difference between the modified cohort component projection and the official population estimate geometrically across each month. The resulting series in red, or "Controlled CCM," bridges the difference between the population estimates implied by the synthetic components of change and those produced by the U.S. Census Bureau to form a consistent time series.

Panel C further adjusts the POP16 series to account for the differences between the Vintage 2020 postcensal estimates derived from the 2010 enumeration and the blended population base adopted with the Vintage 2021 data. This process, shown in orange as the "Intercensal

CCM" series, reconciles the error of closure between the April 1st 2020 estimate from each vintage, again using the Das Gupta (1981) formula. This second adjustment step ensures that the population series is comparable over time without introducing a structural break from the shift in the population base from the Vintage 2021 blended base.

The final step shown in Panel D generates the "Final CNP16" series by subtracting the institutional and resident military population from the resident population, shown in green. This final series further rakes the CNP16 series to match the state CNP16 population total in each month. The final CNP16 series therefore represents a linked population time series that aligns with the CNP16 population universe.

To highlight the resulting resident and civilian noninstitutional populations, I show the monthly population series for four example counties of varying population size in Figure 2. The blue line shows the total population ages 16 plus while the red and orange lines show the civilian noninstitutional population before (red) and after (orange) controlling the series to the official monthly state CNP16 series.

Figure 3 shows the importance of the CNP16 concept, I show two cases where applying the CNP population universe is crucial for estimating an appropriate denominator for labor force statistics: counties with large institutional shares of the overall population and counties with large military bases. Panel A shows the difference between the resident and civilian noninstitutional populations resulting from large institutional populations. A prime example is Crowley County in Colorado, which contains the largest share of prisoners, relative to the total population, of any county in the country (48% as of the 2020 enumeration).

Similarly, counties with large military populations require additional adjustments to match the CNP16 concept. Key examples include two large counties in North Carolina: Cumberland County, which is home to the largest military installation in the world, Fort Bragg, and Onslow County, containing the Marine Corps base Camp Lejeune. In each county, adjusting the resident population to match the CNP16 concept yields sizable differences between the resident population and those likely to be engaged in the civilian labor force. Similar examples include Fort Benning (Chattahoochee County, GA) and Fort Leonard Wood (Pulaski County, MO). For such areas with large military installations, including the sizable military population would inflate the population denominators for traditional labor force measures, thereby underestimating employment to population ratios and labor force participation rates.

5.2 Evaluating Synthetic Components of Change

My first evaluation step compares whether the population estimates computed using the modified cohort component method generally reflect the official population estimates from the U.S. Census Bureau. I first report error metrics between the modified cohort component method and the official U.S. Census Bureau population estimates for ages 16 plus, measured on July 1st of each estimation year. Errors show the relative quality of the estimated components of change in the proposed method compared to the administrative components of change used to produce the official population estimates. Larger errors between the proposed method and the official estimates would indicate that the estimated components of change fail to represent those derived from administrative records. I compute error measures for each estimates year from July 1st 2010 through July 1st 2021 on a rolling basis. I do not include the error of closure between the Vintage 2020 and Vintage 2021 estimates in the error calculations, since the population base for each series differs.

5.2.1 Summary Error Measures

Figure 4 shows the summary error measures that compare the modified cohort component method and the official PEP data itemized by county population size from the 2010 population base. Panel A shows the magnitude differences (MAPE) between the proposed method and the official data are each less than half a percent, demonstrating that the proposed method produces estimates that are very close to the official data produced by the U.S. Census Bureau. Areas with smaller populations, particular in the less than 2,500 and 2,500 to 4,999 ranges shows the largest percent errors of around 0.71% and 0.38%, respectively. Measures of bias (MALPE) in Panel B are similarly small, all larger than -0.25%. Again, the largest bias measures are concentrated in areas with populations below 5,000 residents.

Taken together, the modified cohort component method performs well as compared to the official resident population estimates from PEP. As expected, areas with smaller populations showed larger measures of error and bias compared to larger areas, likely since larger areas tend to have more complete demographic and vital data compared to smaller areas (Rayer, 2008; Wilson et al., 2022).

Comparing the error measures over time, Table 1 summarizes the projection errors of the resident population ages 16 plus by estimation range. The estimation ranges include July to June, April to June in the census years 2010 and 2020, and July to March for 2019 to 2020. Over time, the modified cohort component method produced relatively consistent error ranges with a noticeable spike leading up the COVID-19 pandemic months starting March 2020. A likely explanation for the larger error measures over July 2019 through March 20202 is the impact of the COVID-19 Pandemic on the input data, particularly excess deaths from the pandemic and changes in migration patterns (Gonzalez-Leonardo & Spijker, 2022).

In all cases, the Mean and median absolute percent errors were less than 1% for each estimation range, showing that the proposed method compares favorably to the official data across time. Positive mean and median ALPEs for most estimation ranges show that the estimated components of change tend to overshoot the official estimates, except for an undershot the final estimation range spanning July 2020 through June 2021.

5.2.2 Errors by Geography

Table 2 and Table 3 highlight the variability in model performance relative to population size. Looking first at errors by state in Table 2, I find the largest percentage errors concentrated in lower density states. Specifically, I find the largest errors in Alaska (0.45%), as measured by the MAPE, followed by North Dakota, Nebraska, and South Dakota (each 0.4%). Bias measures were similarly low — less than 0.5% for each state — showing that the estimated components of change well approximate those used by Census Bureau's ADREC method.

Comparatively, mostly positive MALPEs indicate that the estimated components of change tend to overshoot the official population estimates within a tight margin.

Looking further at smaller areas by population size, Table 3 presents the best and worst model fits based on MAPE by county. As expected, the smallest counties showed the largest differences between the proposed method and the official Census Bureau estimates. Relatively large and dense urban counties in core metropolitan areas shows the lowest bias as measured by MAPE, such as Sacramento County, Harris County, and Palm Beach County. Counties with smaller populations tend to have larger projection errors, such as Loving County, TX, the least populous county in the U.S., with a MAPE of 6.4%. In these cases, data on vital statistics is much sparser, owning to issues with data disclosure in the public use data or fewer vital events.

Overall, the proposed modified cohort component method performs well in matching the official population estimates produced by the Census Bureau for the population ages 16 plus. Small error measures for each estimation period show a general confluence between the cumulative components of change used through the Census Bureau's ADREC method and the proposed cohort component method. Accordingly, I conclude that the intra-year variation implied by the vital statistics reflect appropriate seasonal trends in population for areas.

5.3 Evaluating Short-Term Forecasts

Finally, I apply the proposed modified cohort component method to create monthly, short-term population projections. These projections rely on provisional data on vital events from the NCHS, recent population data from the U.S. Census Bureau, and basic assumptions about future migration patterns. As critical inputs to federal statistics production, such as labor force estimates from the Local Area Unemployment Statistics (LAUS) program, I benchmark the modified cohort component estimates against basic extrapolation methods (Rayer, 2008; Wilson, 2022). For a baseline comparison, I forecast CNP16 using linear extrapolation when the change over the past five years was positive and exponential extrapolation if it was negative (Wilson, 2022).

Producing the county-level projections involves applying the modified cohort component model in 9 forward to project the resident population ages 16 plus and 10 and 11 to project CNP16. These projections incorporate the most recent data inputs on aging and net migration rates from the most recent Vintage 2021 data and provisional mortality data from the NCHS. Each projection adds the number of 15-year-olds aging into the 16 plus population each month until July 2022, after which I use the 14-year-olds from the same vintage.

Figure 5 displays the projected and forecast population for counties of varying size starting in July 2021 and running through March 2023 and denoted with dashed lines. Generally, the modified cohort component projections align directionally with forecasts derived from simpler extrapolative methods. A key advantage of the monthly cohort component method is the seasonal variability it produces via its inclusion of more recent vital statistics. Two examples from the Midwest are DeKalb County in Indiana and Douglas County in Illinois.

While both counties are growing steadily in population, the projections computed using cohort components show a more nuanced growth pattern than a simple linear extrapolation. In the case of Douglas County, the population trajectory implied by the vital statistics shows that the county is growing at a slower pace than suggested by a simpler projection method.

Some counties also exhibit time series variation in their population implied by their vital statistics and migration patterns. Producing population projections for these areas would therefore require a more sophisticated model to produce more reasonable projections over time. One example is Etowah County in Alabama. Comparing the simple forecasting method with the modified cohort component method shows a similar downward trend in population; however, the cohort component projections characterize a substantial dip in the population resulting from deaths in the 16 plus population offsetting aging and net migration. In these situations, the modified cohort component method outperforms simpler forecasting methods by incorporating the month-to-month variability in the population attributable to the underlying vital events.

Generally, the modified cohort component method provides more nuanced monthly population projections by incorporating data from the underlying demographic components of change. Comparing the method to standard extrapolative methods, such as linear and exponential models, shows that the modified cohort component method provides more reasonable projections of the population's growth pattern and provides valuable month-to-month variability.

6 Discussion

While monthly labor force measures are widely available for small areas, corresponding information on their population denominators are unavailable. This research develops a unified methodology to estimate the civilian noninstitutional population for counties and equivalents on a monthly basis. The method relies on a modified cohort component method to estimate the monthly resident population and an institutional prevalence rate method to subtract out the institutionalized and resident military population. The method relies on publicly available population data from the U.S. Census Bureau and vital statistics from the National Center for Health Statistics to synthesize monthly demographic components of change and the institutional and military group quarters population. Applying the synthetic components of change to official U.S. Census Bureau estimates each year shows that the method performs well according to MAPE and MALPE measures of fit across a variety of geographies. Further, the method extends to short-term population projections that outperform simpler methods like simple linear extrapolation. The research data provide a novel data set for researchers and planners to study local population and labor market dynamics on a scale that matches the population universe used to produce labor market statistics. Further, the projection method is useful for planners and analysts who require more current population estimates and labor force indicators. Future research would benefit from evaluating more sophisticated methods to address limitations in the mortality and migration data, such modeling suppressed mortality data or alternative measures of local migration.

7 References

Alaska Department of Labor. (2021). *Methods for the alaska population estimates (vintage 2020)*. Alaska Department of Labor and Workforce Development. https://live.laborstats.alaska.gov/pop/estimates/data/AKPopEstMethods.pdf

Beers, H. S. (1945). Six-term formulas for routine actuarial interpolation. *The Record of the American Institute of Actuaries*, *33*(2), 245–260.

Bryan, T. (2004). Population estimates. In J. S. Siegel & D. A. Swanson (Eds.), *The Materials and Methods of Demography* (2nd ed.). Elsevier.

Das Gupta, P. (1981). *Intercensal Estimates for the States [Memorandum]*. U.S. Census Bureau.

DeWaard, J., Hauer, M., Fussell, E., Curtis, K. J., Whitaker, S. D., McConnell, K., Price, K., Egan-Robertson, D., Soto, M., & Castro, C. A. (2022). User Beware: Concerning Findings from the Post 2011–2012 U.S. Internal Revenue Service Migration Data. *Population Research and Policy Review*, *41*(2), 437–448. https://doi.org/10.1007/s11113-021-09663-

Feinstein, C. A. (2002). Seasonality of Deaths in the U.S. by Age and Cause. *Demographic Research*, *6*(2), 469–486. https://doi.org/10.4054/DemRes.2002.6.17

Galvez, J., & McLarty, C. (1996). Measurement of Florida Temporary Residents using a Telephone Survey. *Journal of Economic and Social Measurement*, *22*(1), 25–42. https://doi.org/10.3233/JEM-1996-22102

Gonzalez-Leonardo, M., & Spijker, J. (2022). The Impact of Covid-19 on Demographic Components in Spain, 2020-31: A Scenario Approach. *Population Studies*, *0*(0), 1–17. https://doi.org/10.1080/00324728.2022.2138521

Hamilton, C. H., & Perry, J. (1962). A short method for projecting population by age from one decennial census to another. *Social Forces*, *41*(2), 163–170. https://doi.org/10.2307/2573607

Hauer, M. E. (2019). Population Projections for U.S. Counties by Age, Sex, and Race Controlled to Shared Socioeconomic Pathway. *Scientific Data*, *6*, 190005. https://doi.org/10.1038/sdata.2019.5

Lam, D. A., & Miron, J. A. (1996). The Effects of Temperature on Human Fertility. *Demography*, *33*(3), 291–305. https://doi.org/10.2307/2061762

Land, K. C., & Hough, G. C. (1986). Improving the accuracy of intercensal estimates and postcensal projections of the civilian noninstitutional population. *Journal of the American Statistical Association*, *81*(393), 62–74. https://doi.org/10.1080/01621459.1986.10478238

McHugh, K. E., Hogan, T. D., & Happel, S. K. (1995). Multiple residence and cyclical migration: A life course perspective. *The Professional Geographer*, *47*(3), 251–267. https://doi.org/10.1111/j.0033-0124.1995.251/_q.x

Osterman, M. J. K., Hamilton, B. E., Martin, J. A., Driscoll, A. K., & Valenzuela, C. P. (2023). Births: Final data for 2021. *National Vital Statistics Reports*, *72*(1). https://doi.org/10.15620/cdc:122047

Paglino, E., Lundberg, D. J., Zhou, Z., Wasserman, J. A., Raquib, R., Luck, A. N., Hempstead, K., Bor, J., Preston, S. H., Elo, I. T., & Stokes, A. C. (2023). Monthly excess mortality across counties in the united states during the COVID-19 pandemic, march 2020 to february 2022. *Science Advances*, *9*(25), eadf9742. https://doi.org/10.1126/

sciadv.adf9742

Perz, S. G. (2004). Population change. In J. S. Siegel & D. A. Swanson (Eds.), *The Materials and Methods of Demography* (2nd ed.). Elsevier.

Preston, S., Heuveline, P., & Guillot, M. (2000). *Demography: Measuring and Modeling Population Processes*. Wiley-Blackwell.

Rayer, S. (2008). Population Forecast Errors: A Primer for Planners. *Journal of Planning Education and Research*, *27*(4), 417–430. https://doi.org/10.1177/0739456X07313925

Rosenberg, H. M. (1965). Seasonal adjustment of vital statistics by electronic computer. *Public Health Reports*, *80*(3), 201–210. https://pubmed.ncbi.nlm.nih.gov/14265432/

Smith, S. K. (1989). Toward a methodology for estimating temporary residents. *Journal of the American Statistical Association*, *84*(406), 430–436. https://doi.org/10.1080/01621459.1989.10478787

Smith, S. K., & Mandell, M. (1984). A comparison of population estimation methods: Housing unit versus component II, ratio correlation, and administrative records. *Journal of the American Statistical Association*, *79*(386), 282–289. https://doi.org/10.1080/01621459.1984.10478042

Smith, S. K., & Sincich, T. (1990). The relationship between the length of the base period and population forecast errors. *Journal of the American Statistical Association*, *85*(410), 367–375. https://doi.org/10.2307/2289772

State of California, Department of Finance. (2023). *E-1 population estimates for cities, counties and the state with annual percent change — january 1, 2022 and 2023*. State of California, Department of Finance. https://dof.ca.gov/forecasting/demographics/estimates-e1/

U.S. Bureau of Economic Analysis. (2022). *Note on Per Capita Personal Income and Population*. U.S. Bureau of Economic Analysis. https://www.bea.gov/note-capita-personal-income

U.S. Bureau of Labor Statistics. (2022). *Current Population Survey*. U.S. Bureau of Labor Statistics. https://www.bls.gov/opub/hom/cps/home.htm

U.S. Census Bureau. (2021a). *Annual Resident Population Estimates, Estimated Components of Resident Population Change, and Rates of the Components of Resident Population Change for States and Counties: April 1, 2010 to July 1, 2020 [dataset]*. U.S. Census Bureau. https://www2.census.gov/programs-surveys/popest/technical-documentation/file-layouts/2010-2020/co-est2020-alldata.pdf

U.S. Census Bureau. (2002). *Current population survey: Design and methodology*. U.S. Census Bureau; Technical Paper 63RV.

U.S. Census Bureau. (2009). *Methodology for the State and County Total Resident Population Estimates (Vintage 2009): April 1, 2000 to July 1, 2009*. U.S. Census Bureau.

U.S. Census Bureau. (2012a). *2010 census summary file 1*. U.S. Census Bureau; 2010 Census of Population and Housing.

U.S. Census Bureau. (2012b). *Methodology for the Intercensal Population and Housing Unit Estimates: 2000 to 2010*. U.S. Census Bureau. https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/intercensal/2000-2010-intercensal-estimates-methodology.pdf

U.S. Census Bureau. (2021b). *Methodology for the United States Population Estimates: Vintage 2021*. U.S. Census Bureau.

U.S. Census Bureau. (2022a). *Annual County and Puerto Rico Municipio Resident Population Estimates by Selected Age Groups and Sex: April 1, 2010 to July 1, 2019; April 1, 2020; and July 1, 2020 [dataset].* https://www.census.gov/programs-surveys/popest/technical-documentation/research/evaluation-estimates/2020-evaluation-estimates/2010s-counties-total.html

U.S. Census Bureau. (2022b). *Annual County and Puerto Rico Municipio Resident Population Estimates by Selected Age Groups and Sex: April 1, 2020 to July 1, 2021 [dataset].* https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html

U.S. Census Bureau. (2022c). *Annual Resident Population Estimates, Estimated Components of Resident Population Change, and Rates of the Components of Resident Population Change for States and Counties: April 1, 2020 to July 1, 2021 [dataset].* U.S. Census Bureau. https://www2.census.gov/programs-surveys/popest/technical-documentation/file-layouts/2020-2021/CO-EST2021-ALLDATA.pdf

Wilson, T. (2022). Preparing Local Area Population Forecasts Using a Bi-Regional Cohort-Component Model Without the Need for Local Migration Data. *Demographic Research*, *46*(32), 919–956. https://doi.org/10.4054/DemRes.2022.46.32

Wilson, T., Grossman, I., Alexander, Monica, Rees, P., & Temple, J. (2022). Methods for Small Area Population Forecasts: State-of-the-Art and Research Needs. *Population Research and Policy Review*, *41*(3), 865–898. https://doi.org/10.1007/s11113-021-09671-6

Xu, J., Murphy, S. L., Kochanek, K. D., & Arias, E. (2022). *Mortality in the united states, 2021.* NCHS Data Brief No. 456. https://doi.org/10.15620/cdc:122516

8 Figures

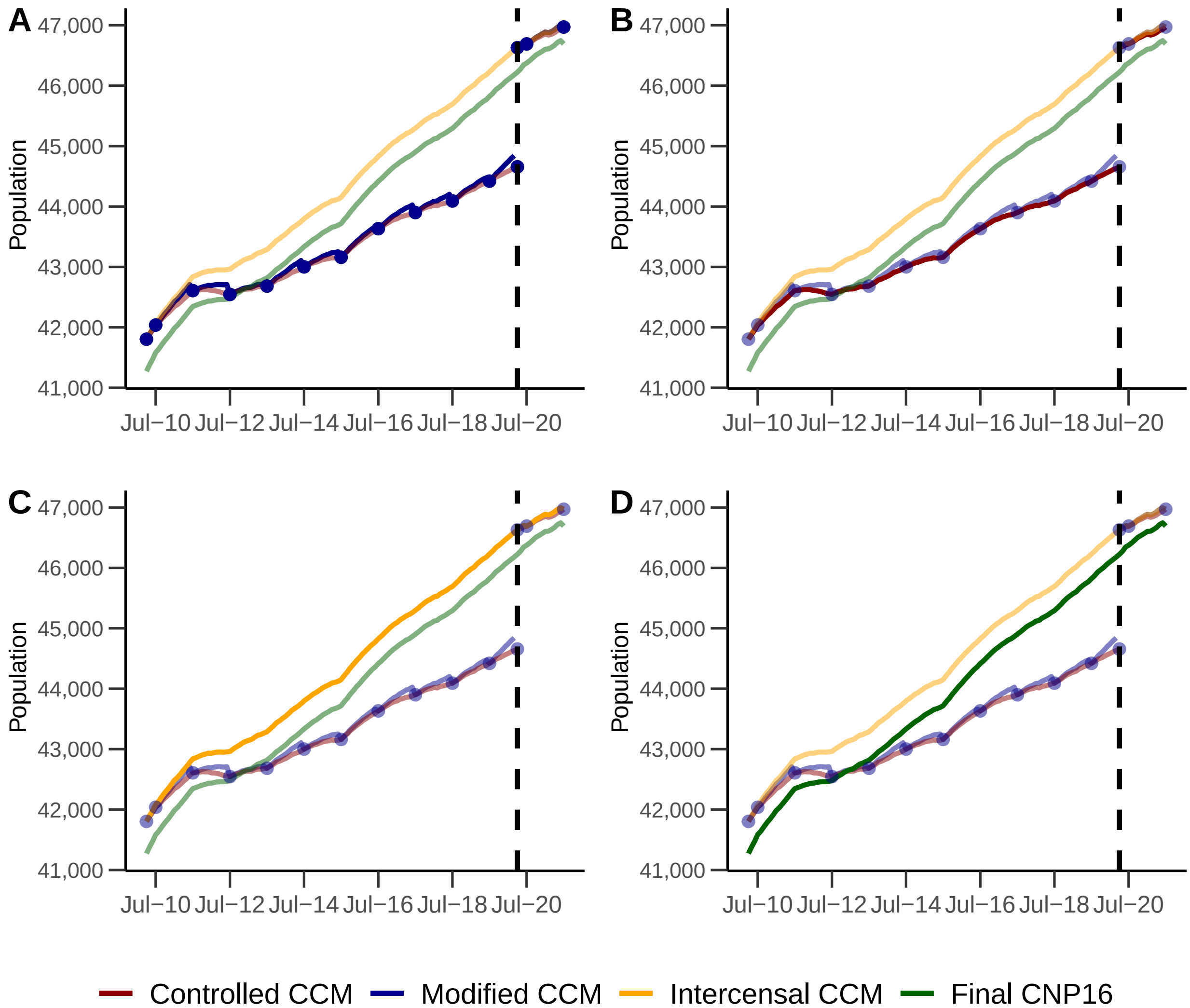

Figure 1: Estimation Stages for the Resident and Civilian Noninstiutional Population Ages 16 Plus

Notes: Figure shows each step of the estimation process for POP16 and CNP16. The vertical dashed line indicates the shift from the Vintage 2020 data to Vintage 2021. Dots indicate the official population estimates from the Census Bureau. Panel A shows the synthetic components of change applied each year. Panel B controls the modified CCM to the official estimates. Panel C bridges the Vintage 2020 and Vintage 2021 data. Panel D subtracts the noninstitutional and military quarters populations.

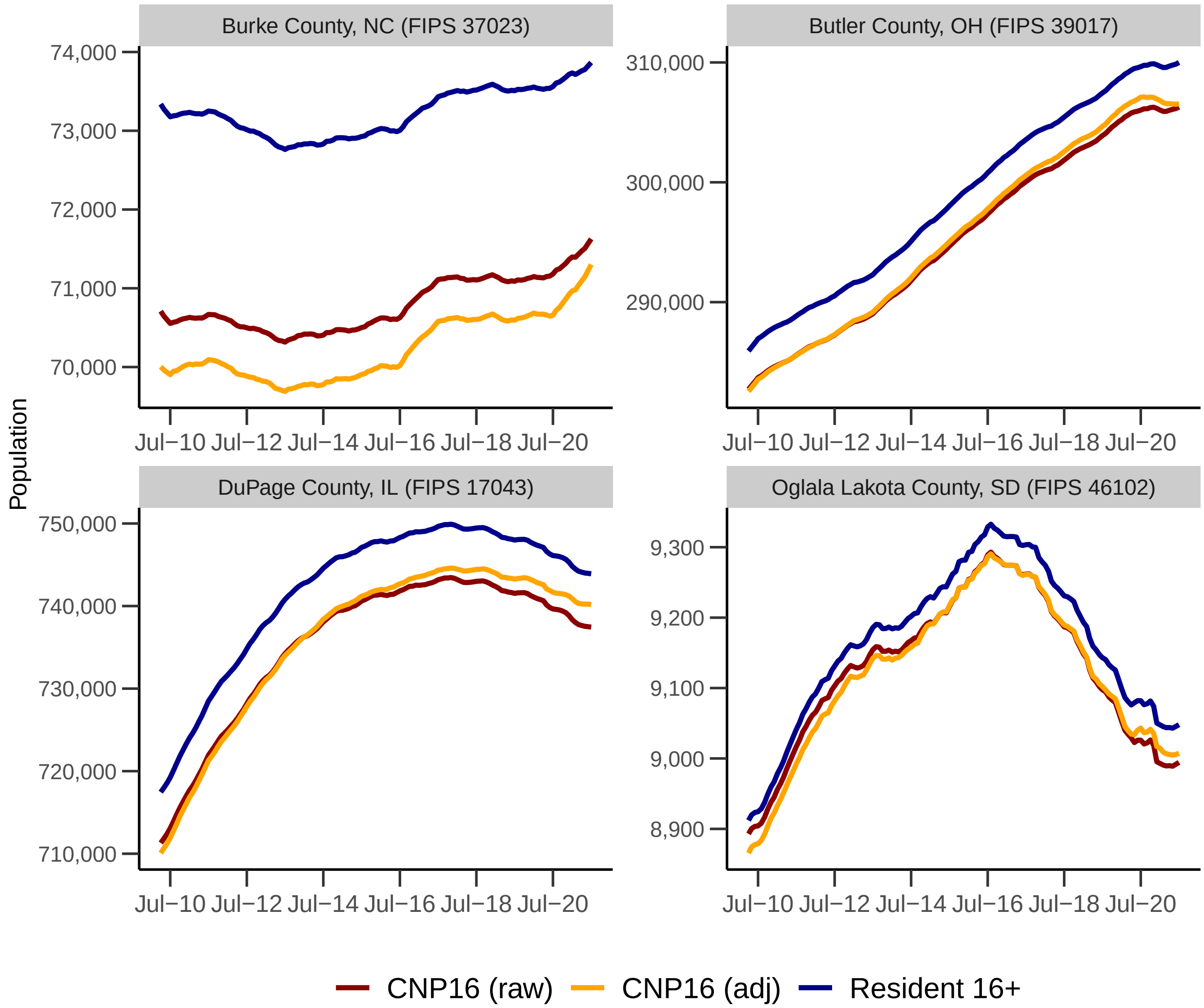

Figure 2: Resident and Civilian Noninstitutional Population for Selected Counties

Notes: Figure shows the total resident and civilian noninstitutional population ages 16 plus for selected counties of varying population size. Final CNP16 estimates are controlled to the monthly state CNP16 series published by the U.S. Bureau of Labor Statistics.

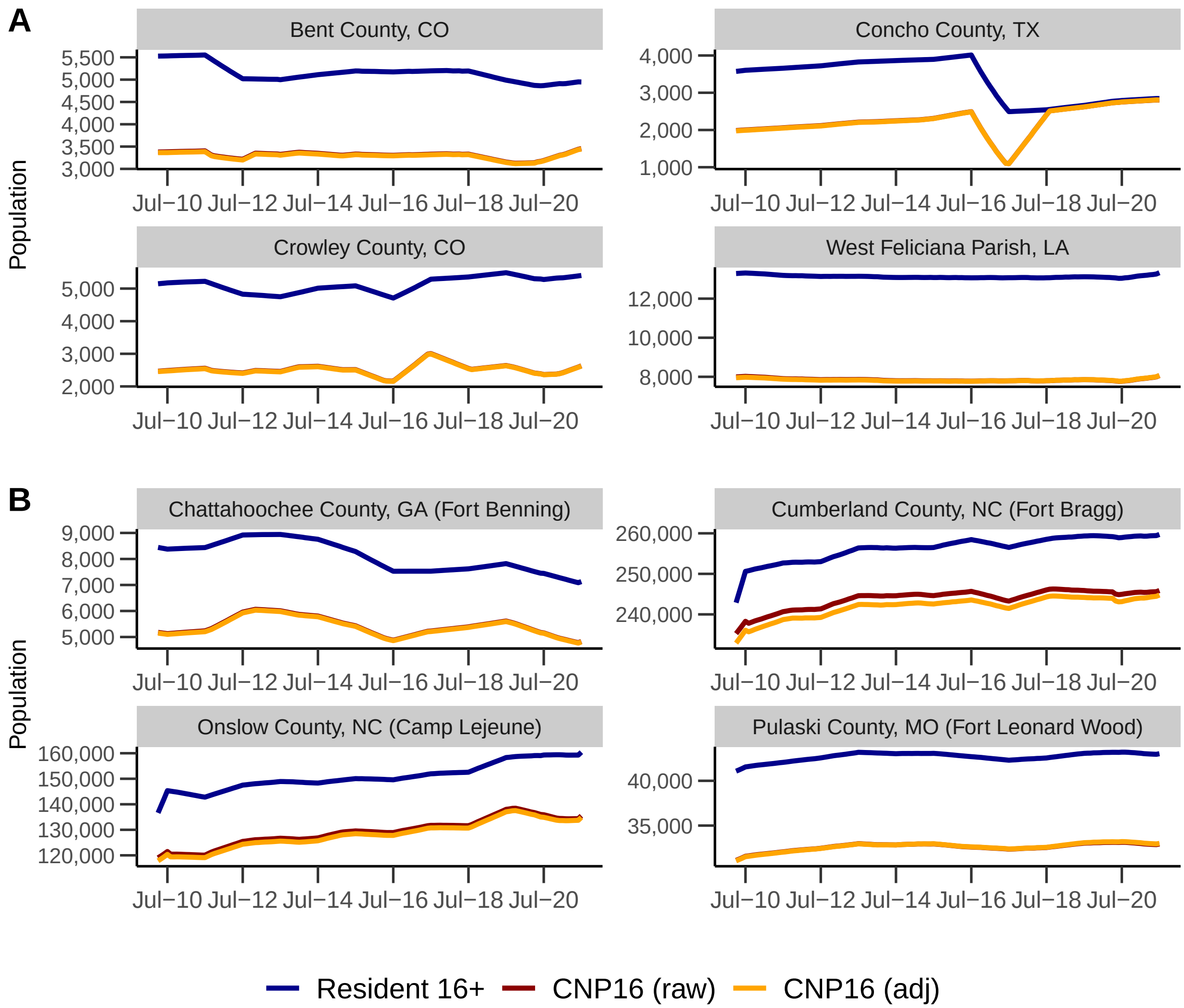

Figure 3: Resident and Civilian Noninstitutional Population for Counties with Large Prisons and Military Bases

Notes: Figure shows the total resident and civilian noninstitutional population ages 16 plus for counties with large prison shares of the population and large military bases. Final CNP16 estimates are controlled to the monthly state CNP16 series published by the U.S. Bureau of Labor Statistics

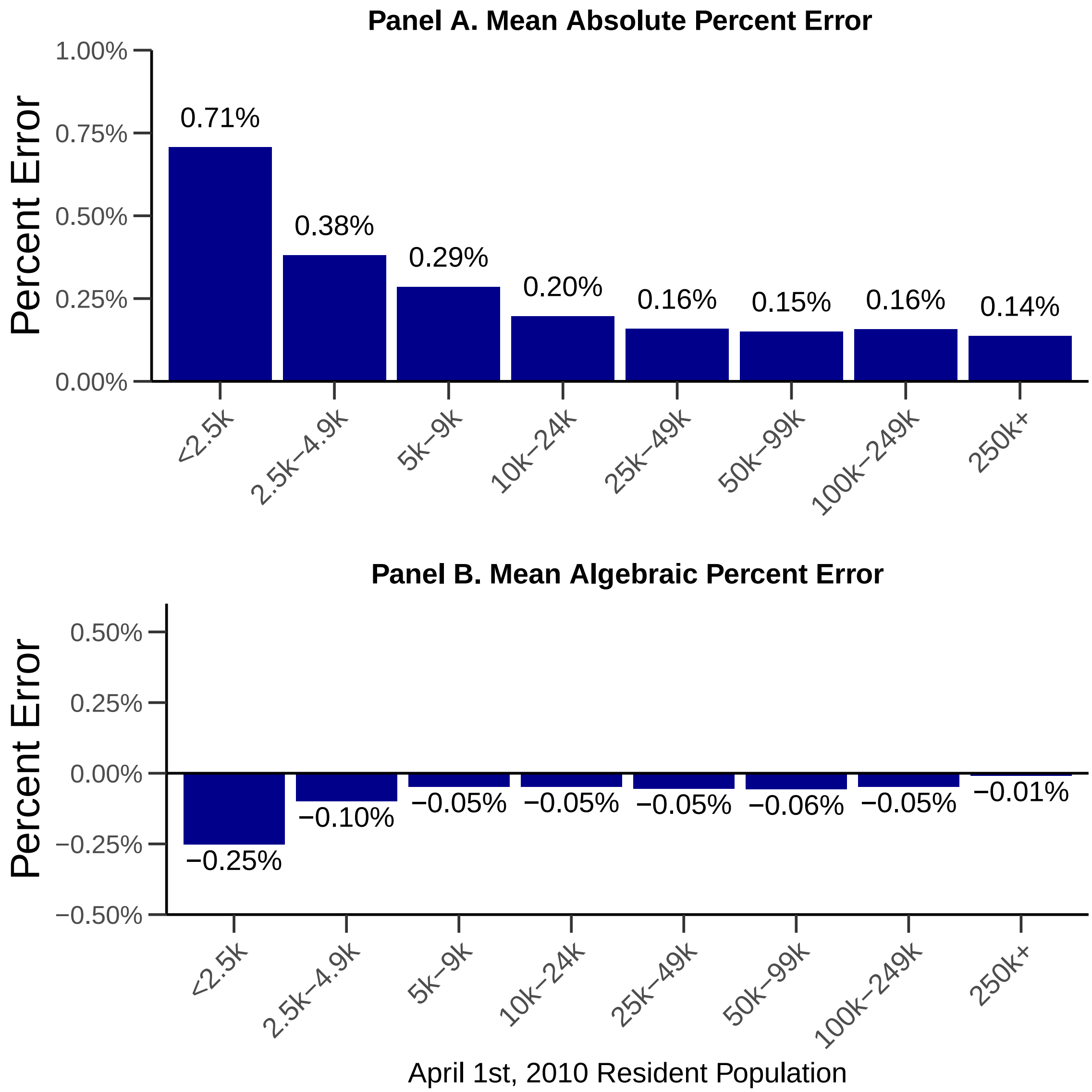

Figure 4: MAPE and MALPE by County Population Size

Notes: Figure shows mean and median error measures that compare the difference between the modified cohort component model and the official Census Bureau PEP estimates. MAPE denotes the Mean Absolute Percent Error and MALPE denotes the Mean Algebraic Percent Error. Population size is based on areas' April 2010 population base. Errors measures to not include the error of closure between the Vintage 2020 and 2021 data.

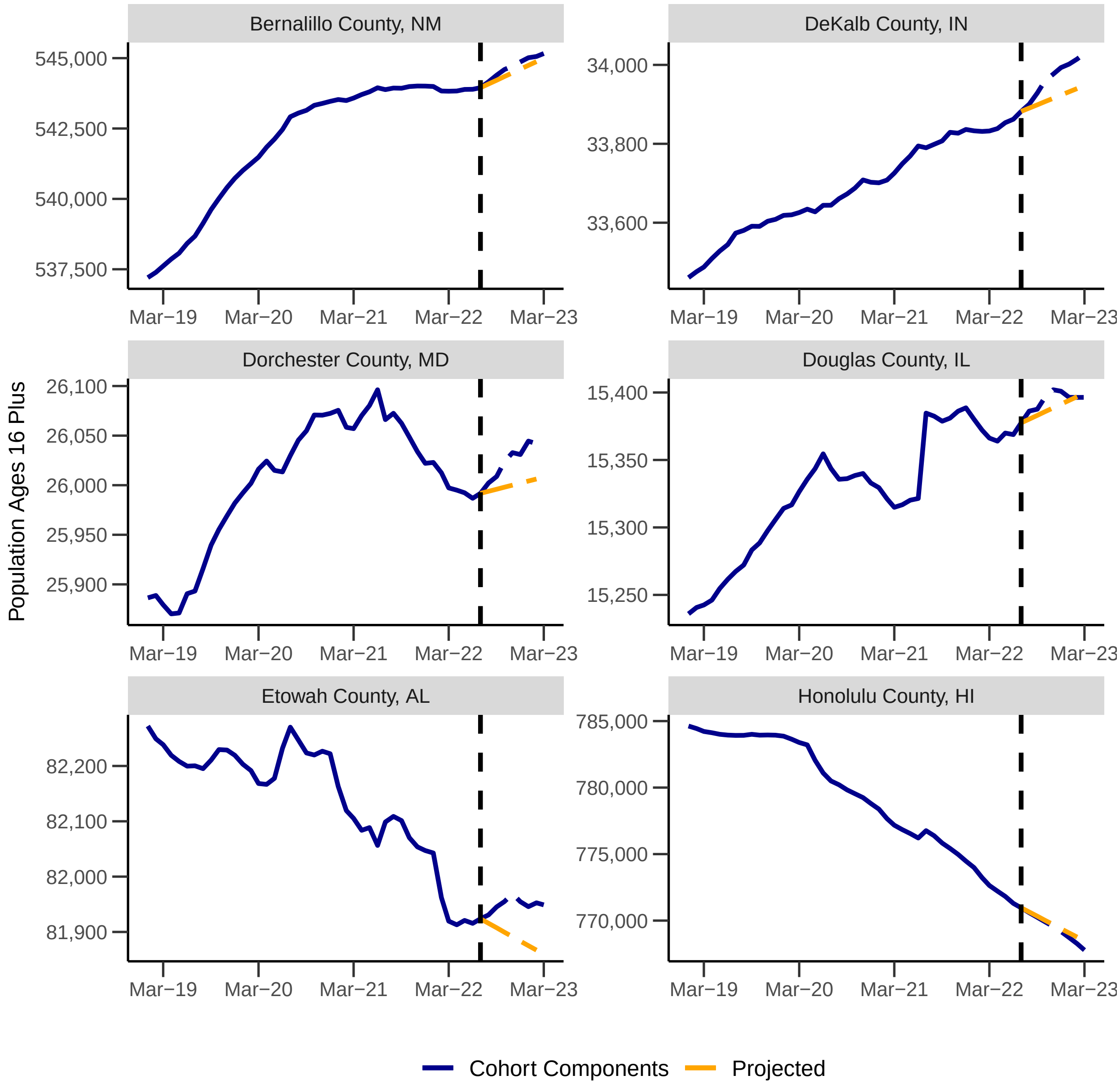

Figure 5: CNP16 Projections, Modified Cohort Components and Simple Forecasts

Notes: Figure shows projections using the proposed cohort component method against simple forecasting methods. Forecasts follow Wilson (2022) and use a linear model when population growth for the past 5 years was positive and an exponential model when it was negative.

9 Tables

Table 1: Errors by Estimation Period

Estimation Range	Mean APE	Mean ALPE	Median APE	Median ALPE
Apr-10 - Jun-10	0.11%	-0.01%	0.08%	-0.01%
Jul-10 - Jun-11	0.23%	-0.09%	0.15%	-0.07%
Jul-11 - Jun-12	0.24%	-0.07%	0.15%	-0.06%
Jul-12 - Jun-13	0.25%	-0.10%	0.17%	-0.08%
Jul-13 - Jun-14	0.24%	-0.09%	0.17%	-0.08%
Jul-14 - Jun-15	0.25%	-0.10%	0.17%	-0.09%
Jul-15 - Jun-16	0.25%	-0.11%	0.17%	-0.10%
Jul-16 - Jun-17	0.25%	-0.11%	0.17%	-0.10%
Jul-17 - Jun-18	0.23%	-0.09%	0.16%	-0.08%
Jul-18 - Jun-19	0.24%	-0.01%	0.16%	-0.02%
Jul-19 - Mar-20	0.34%	-0.01%	0.23%	-0.02%
Apr-20 - Jun-20	0.11%	-0.05%	0.08%	-0.06%
Jul-20 - Jun-21	0.29%	0.03%	0.20%	0.02%

Notes: Table shows mean and median error measures that compare the difference between the resident population ages 16 plus produces using the modified cohort component method and the official Census Bureau PEP estimates. APE denotes the Absolute Percent Error and ALPE denotes the Algebraic Percent Error.

Table 2: Model Error Measures by State, MAPE and MALPE

State Abbv	State Name	Mean APE	Mean ALPE	Median APE	Median ALPE
AK	Alaska	0.44%	-0.16%	0.27%	-0.07%
AL	Alabama	0.15%	-0.02%	0.11%	-0.02%
AR	Arkansas	0.18%	-0.01%	0.14%	-0.02%
AZ	Arizona	0.19%	0.08%	0.15%	0.08%
CA	California	0.17%	-0.06%	0.13%	-0.06%
CO	Colorado	0.28%	-0.04%	0.17%	-0.01%
CT	Connecticut	0.13%	-0.12%	0.12%	-0.12%
DC	District of Columbia	0.33%	0.32%	0.34%	0.34%
DE	Delaware	0.13%	0.05%	0.09%	0.04%
FL	Florida	0.18%	-0.02%	0.12%	0.00%
GA	Georgia	0.27%	0.02%	0.18%	0.00%
HI	Hawaii	0.20%	0.01%	0.08%	0.00%
IA	Iowa	0.22%	-0.12%	0.17%	-0.12%
ID	Idaho	0.34%	-0.11%	0.22%	-0.10%
IL	Illinois	0.19%	-0.07%	0.14%	-0.07%
IN	Indiana	0.17%	-0.11%	0.14%	-0.10%
KS	Kansas	0.34%	-0.07%	0.22%	-0.06%
KY	Kentucky	0.20%	-0.03%	0.14%	-0.04%
LA	Louisiana	0.19%	0.05%	0.13%	0.03%
MA	Massachusetts	0.13%	-0.03%	0.12%	-0.05%
MD	Maryland	0.19%	-0.10%	0.13%	-0.09%
ME	Maine	0.11%	-0.08%	0.08%	-0.06%
MI	Michigan	0.14%	-0.08%	0.11%	-0.08%
MN	Minnesota	0.21%	-0.13%	0.16%	-0.12%
MO	Missouri	0.20%	-0.05%	0.15%	-0.05%
MS	Mississippi	0.21%	0.09%	0.15%	0.06%
MT	Montana	0.39%	-0.13%	0.24%	-0.08%
NC	North Carolina	0.18%	-0.02%	0.11%	-0.04%
ND	North Dakota	0.40%	-0.17%	0.28%	-0.13%
NE	Nebraska	0.41%	-0.22%	0.28%	-0.15%
NH	New Hampshire	0.12%	-0.08%	0.10%	-0.08%
NJ	New Jersey	0.16%	-0.10%	0.13%	-0.10%
NM	New Mexico	0.25%	0.01%	0.16%	0.00%
NV	Nevada	0.29%	-0.02%	0.16%	-0.02%
NY	New York	0.14%	-0.06%	0.11%	-0.08%
OH	Ohio	0.14%	-0.10%	0.12%	-0.10%
OK	Oklahoma	0.24%	-0.08%	0.17%	-0.06%
OR	Oregon	0.18%	-0.06%	0.13%	-0.05%
PA	Pennsylvania	0.14%	-0.07%	0.11%	-0.09%
RI	Rhode Island	0.14%	-0.09%	0.11%	-0.05%

State Abbv	State Name	Mean APE	Mean ALPE	Median APE	Median ALPE
SC	South Carolina	0.16%	0.03%	0.11%	0.01%
SD	South Dakota	0.38%	-0.11%	0.26%	-0.09%
TN	Tennessee	0.17%	-0.03%	0.11%	-0.05%
TX	Texas	0.35%	-0.12%	0.20%	-0.07%
UT	Utah	0.31%	-0.12%	0.21%	-0.07%
VA	Virginia	0.23%	-0.02%	0.17%	-0.04%
VT	Vermont	0.13%	-0.08%	0.11%	-0.07%
WA	Washington	0.18%	-0.08%	0.12%	-0.06%
WI	Wisconsin	0.16%	-0.09%	0.12%	-0.09%
WV	West Virginia	0.16%	-0.03%	0.12%	-0.03%
WY	Wyoming	0.24%	-0.04%	0.17%	-0.06%

Alaska, North Dakota, Nebraska

Notes: Table shows mean and median error measures that compare the difference between the modified cohort component model and the official Census Bureau PEP estimates. APE denotes the Absolute Percent Error and ALPE denotes the Algebraic Percent Error.

Table 3: Average and Median Errors for Counties

FIPS	Area Name	POP16	MAPE	MALPE	Med. APE	Med. ALPE
06067	Sacramento County	1,261,497	0.02%	0.00%	0.02%	0.01%
41039	Lane County	322,848	0.03%	-0.02%	0.01%	0.00%
25001	Barnstable County	203,459	0.03%	0.00%	0.02%	-0.01%
17113	McLean County	139,014	0.03%	-0.01%	0.02%	-0.01%
48201	Harris County	3,626,621	0.03%	0.02%	0.02%	0.02%
42071	Lancaster County	439,163	0.04%	0.00%	0.05%	0.01%
12099	Palm Beach County	1,247,763	0.04%	-0.02%	0.02%	-0.01%
26049	Genesee County	324,813	0.04%	-0.03%	0.04%	-0.03%
10003	New Castle County	464,715	0.04%	-0.03%	0.04%	-0.03%
42003	Allegheny County	1,032,339	0.04%	0.03%	0.04%	0.03%
31075	Grant County	454	1.23%	-0.71%	0.94%	-0.75%
02282	Yakutat City and Borough	575	1.24%	-0.88%	0.83%	-0.66%
28055	Issaquena County	1,200	1.25%	1.19%	1.06%	1.06%
31009	Blaine County	372	1.25%	-0.86%	0.58%	-0.46%
46102	Oglala Lakota County	9,048	1.35%	-1.35%	1.33%	-1.33%
48033	Borden County	498	1.40%	-1.17%	0.94%	-0.76%
13053	Chattahoochee County	7,130	1.60%	1.52%	1.05%	1.05%
48269	King County	194	1.68%	-1.29%	1.33%	-1.03%
48261	Kenedy County	286	1.75%	-0.52%	1.51%	-0.19%
48301	Loving County	36	5.87%	-2.73%	5.49%	-0.99%

Notes: Table shows the top and bottom 5 counties ranked by MAPE. Mean and median error measures that compare the difference between the modified cohort component model and the official Census Bureau PEP estimates. APE denotes the Absolute Percent Error and ALPE denotes the Algebraic Percent Error. Resident population age 16+ are as of July 1st, 2021 from the Vintage 2021 Population Estimates Program.

Online Appendix for: Estimating the Civilian Noninstitutional Population for Small Areas

A Modified Cohort Component Approach Using Public Use Data

Andrew C. Forrester*

U.S. Bureau of Labor Statistics

*Forrester: forrester.andrew@bls.gov, Division of Local Area Unemployment Statistics, Office of Employment and Unemployment Statistics, U.S. Bureau of Labor Statistics. All views are my own and do not necessarily reflect those of the U.S. Bureau of Labor Statistics, the Department of Labor, or the United States.

A Comparison with the American Community Survey

One incomplete benchmark for the modified cohort component method is the American Community Survey (ACS), which provides annual, county-level estimates of the CNP16. These data tabulate the CNP16 for a subset of areas through a disability status table, Table S1811. Due to sample size constraints on the annual ACS sample, the 1-year estimates only provide a subset of counties each year, of which a smaller subset of counties have published CNP16 estimates. The ACS 1-year estimates are also produced by weighting respondents using the current year vintage population controls from PEP and are not revised using intercensal adjustments. Consequently, there are a few details to consider when comparing the ACS-derived CNP16 estimates with those produced using cohort components.

A.1 *Caveat Emptor*

Conceptually, each method estimates the same population universe; however, the ACS design and publishing standards complicate any comparison between the two data series. For one, the ACS estimates, by design, conform to independent population control totals developed by the *Population Estimates Program* (PEP) that represent the resident population as of July 1st each year. For years in between the decennial enumerations, the ACS is controlled to the current year postcensal population estimates and only use intercensal estimates in decennial years (U.S. Census Bureau, 2021). For example, the population bases will match for the 2010 census and diverge each year, as the ACS uses Vintage 2011 estimates for the 2011 ACS, Vintage 2012 for the 2012 ACS, and so on. This feature of the ACS design means the ACS estimates will reflect different population controls each year compared to the population estimates used in the modified cohort component method between census years.

The ACS also limits the number of areas published in the 1-year estimates to a population threshold of 65,000 and above. Since the ACS 1-year data only contain CNP16 estimates for around 100 large counties (around 44% of the national CNP16 in the 2021 ACS), county-to-county comparisons will capture the efficacy of the modified cohort component method for only a small fraction of more populous areas where vital statistics are more likely to be complete.

A final caveat to this benchmark is that the 2020 ACS estimates were not published for counties. Due to data collection errors resulting from the COVID-19 pandemic, the U.S. Census Bureau was unable to collect a robust sample for the ACS over the course of the year and instead published only a limited set of state-level estimates, developed using an experimental weighting methodology. I therefore leave out a comparison for 2020 to maintain consistency between the ACS 1-year estimates over time.

A.2 Direct Comparisons

I make two types of comparisons with the ACS — between the percentage and level differences across each method. Specifically, I compare the ACS estimates with July 1st cohort component estimates. Table 1 summarizes each set of comparisons, while Figure 1 and Figure 2 display each comparison visually.

First, I compare the percentage differences (MAPE and MALPE) between the modified cohort component method and the ACS 1-year estimates over time in Figure 1. In each calculation I assume the ACS 1-year estimate is the "true" CNP16 value. Panel A shows that the the percentage differences between the ACS and the modified cohort component method increase almost monotonically over 2010 to 2019, increasing from a low of 0.4% in 2010 to 1.7% in 2019. It is unclear whether the discrepancy between the two series reflects the difference in ACS survey weights or the increasing number of published counties in the ACS data. Supporting the population controls hypothesis, the differences between the ACS and modified cohort component method subsequently drop to a level of 0.4% in 2021, similar in magnitude to the 2010 estimates closer to the 2010 census. Looking at the direction of differences between the ACS and modified cohort component method in Panel B, I find that the modified cohort component method tends to overshoot the ACS estimates in most years. Again, it is unclear whether this results from the aforementioned difference in population controls using in the underlying ACS estimates or the number of published areas represented in the ACS 1-year data.

Second, Figure 2 shows the level differences between the ACS and modified cohort component method across all matched counties with the associated data in Table 1. The modified cohort component method tracks the ACS well in level terms, both for the unadjusted and raked series. In every year, the differences between the total CNP16 were less than 1%, comparing both the adjusted and unadjusted series.

Across all matched counties from the ACS, the modified cohort component method produces estimates that are nearly identical to those produced by the ACS with the appropriate caveats. While percentage errors between the ACS and the cohort component estimates increased each year between 2010 and 2019, the disparity likely arose from the population controls used to weight ACS respondents. Aggregating all published counties each year, the level and percentage differences are minimal and amount to less than 1% each year. Taken together — and with appropriate caution — the ACS and proposed cohort component method produce virtually indistinguishable results.

B Updating Institutional Prevalence Rates

The most recent group quarters data by facility type, age, and sex were released on May 25, 2023 from the 2020 census as part of the Demographic and Housing Characteristics (DHC) file. These data include the enumeration of the population residing in group quarters by facility type, sex, and age group. To compute institutional prevalence rates using the new 2020 data, I first derive the the institutional and resident military populations by 5-year age groups using the 2020 census data in Tables PCO1 (totals), PCO2 (institutional), and PCO9 (military). I recode the under 20 years old age group to reflect the 15 to 19 age range before computing the prevalence rates, since new military recruits must be at least 17 years old by law.[1] Using the standard 5-year age-intervalled data, I apply the Beers (1945) formula to compute the 16 to 19 age group.

The next consideration is how to reconcile the difference between the 2010 and 2020 institutional prevalence rates. Since the group quarters population does not change according to the standard demographic components of change, I linearly interpolate the institutional prevalence rates between April 2010 and April 2020 to prevent a discontinuous break in the series when incorporating the new 2020 data. After April 2020 I follow the literature by carrying the institutional prevalence rate forward through the projection horizon, as the standard is to assume no change in the group quarters population unless other data are received (Bryan, 2004).

B.1 Results

Figure 3 compares the age distribution of the total group quarters and institutional and resident military populations between the 2010 and 2020 enumerations. Each chart compares the share of each age group relative to the respective total, i.e., the 20 to 24 year old share of the institutional and military group quarters population. Moving from 2010 to 2020 I find the largest shifts in the institutional and resident military populations was into the ages 65 plus age group. This trend likely reflects general aging in the U.S. population, as institutional group quarters include nursing facilities and hospices.

Next, I examine the effect of integrating the 2020 enumeration-based institutional prevalence rates into the overall CNP16 estimates. Figure 4 compares three series: the CNP16 series produced with the 2010-based institutional prevalence rates (blue), integrating the 2020 rates as described above (red), and with the final raked estimates to the independent CNP16 series from the U.S. Bureau of Labor Statistics. Both over the historical and projected periods, both methods produce nearly identical estimates, with the 2010-based institutional prevalence rates producing slightly larger estimates at the national level.

[1] 10 U.S.C. 505.(a) requires that new enlistments must be at least 17 years of age.

C Additional Figures

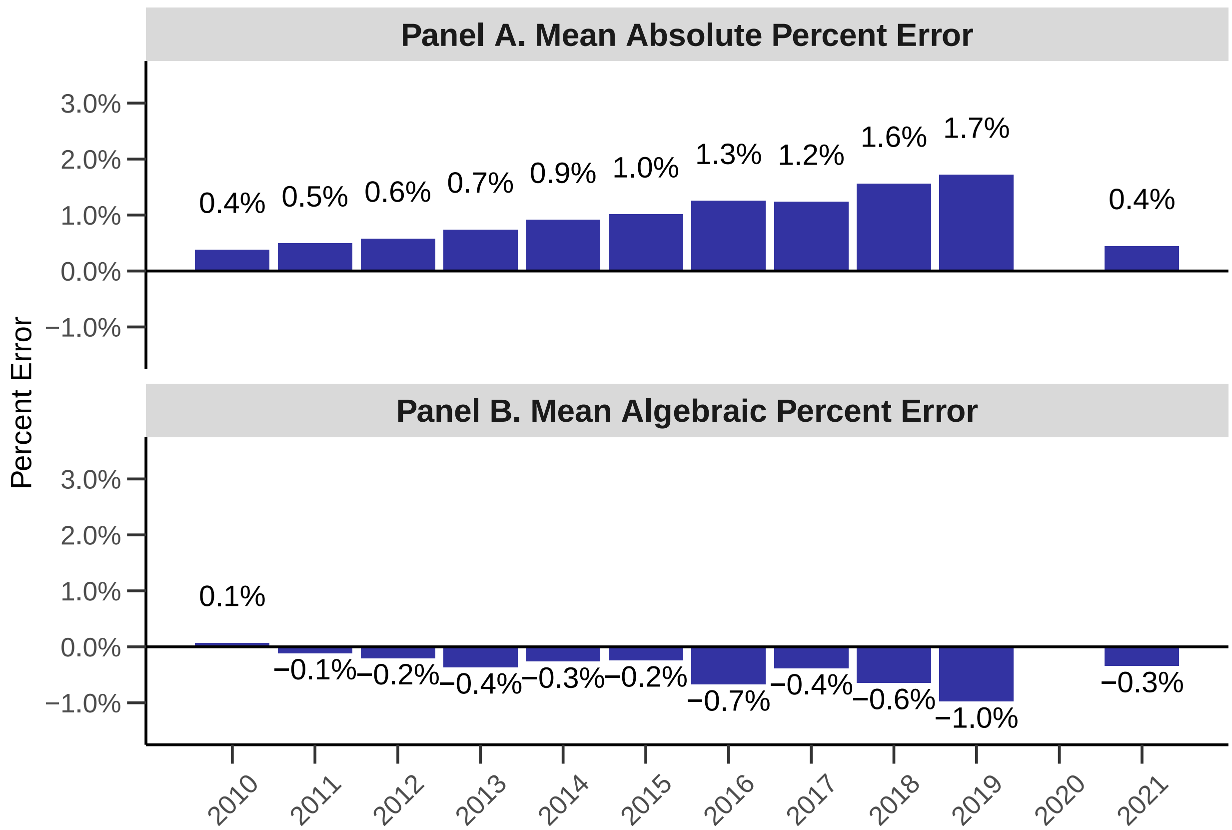

Figure A 1: Percent Errors Between CNP16 from the ACS and Cohort Components

Notes: Figure shows the MAPE and MALPE between the CNP16 from the ACS 1-year Table S1811 and the modified cohort component estimates, raked to the state CNP16 totals from the Current Population Survey (CPS), as of July 1st each year. ACS data are weighted based on the current year vintage population estimates and do not include an intercensal adjustment. 2020 ACS data for counties were not published due to data collection issues resulting from the COVID-19 pandemic.

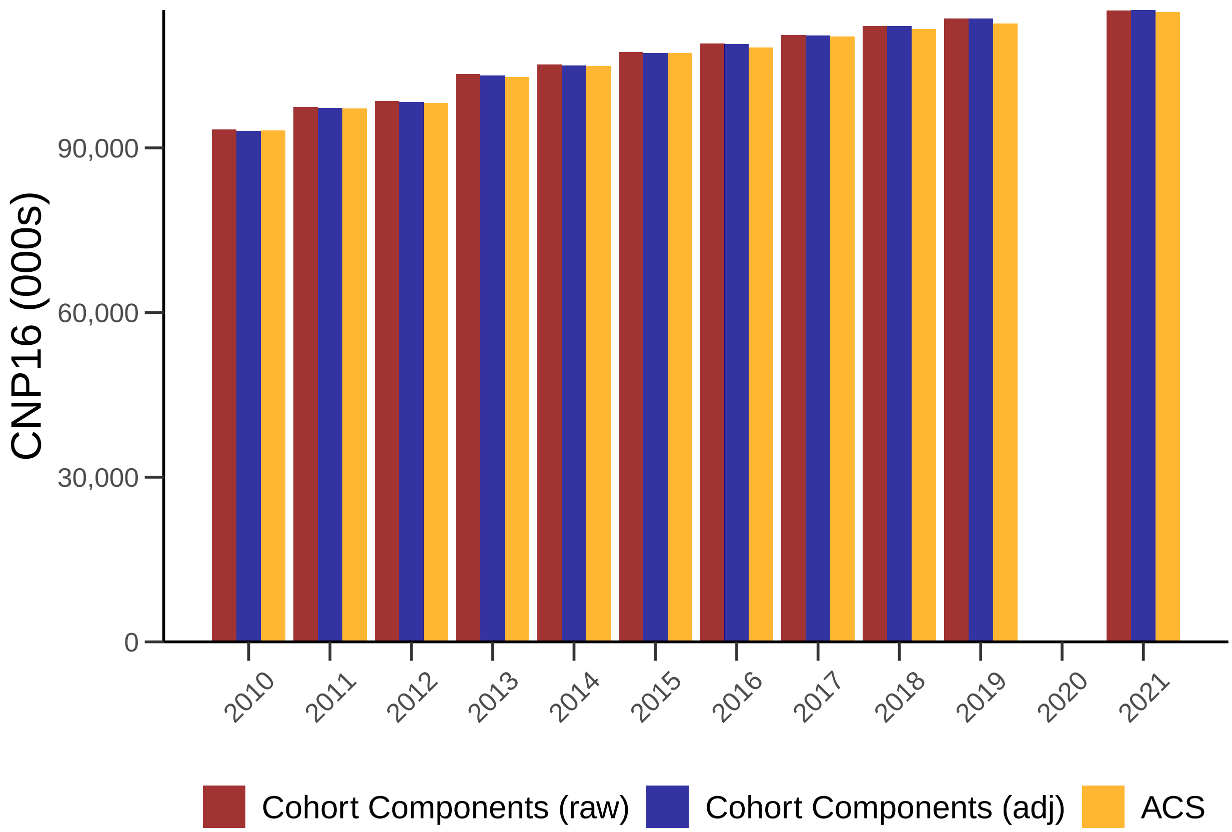

Figure A 2: Level Differences Between CNP16 from the ACS and Cohort Components

Notes: Figure shows the level differences between the CNP16 from the ACS 1-year Table S1811 and the modified cohort component estimates, raked to the state CNP16 totals from the Current Population Survey (CPS), as of July 1st each year. ACS data are weighted based on the current year vintage population estimates and do not include an intercensal adjustment. 2020 ACS data for counties were not published due to data collection issues resulting from the COVID-19 pandemic.

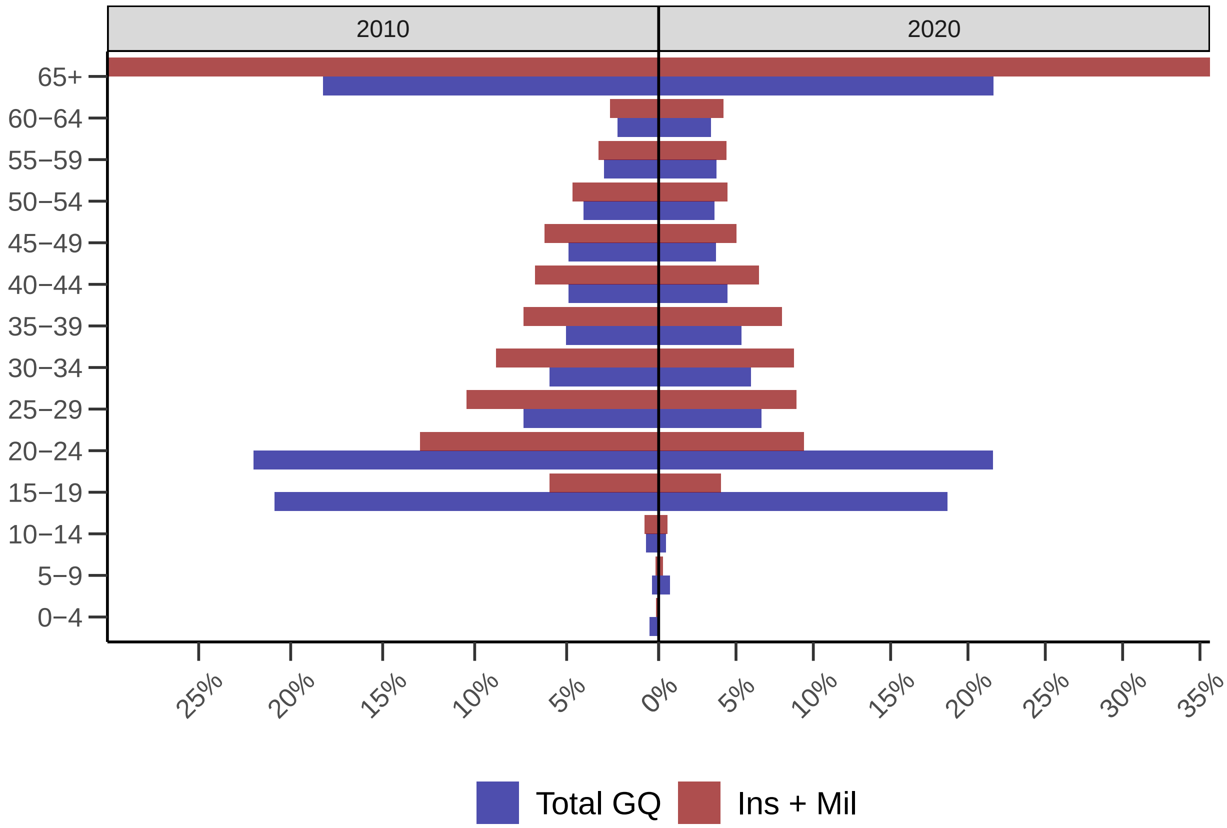

Figure A 3: Group Quarters Population Population Share for the Total and Institutional Plus Resident Military

Notes: Figure shows the share of the total group quarters and institutional and resident military population in each each group in the 2010 and 2020 censuses. Data from the 2010 census are from Summary File 1 and data from the 2020 census are from the Demographic and Housing Characteristics (DHC) file.

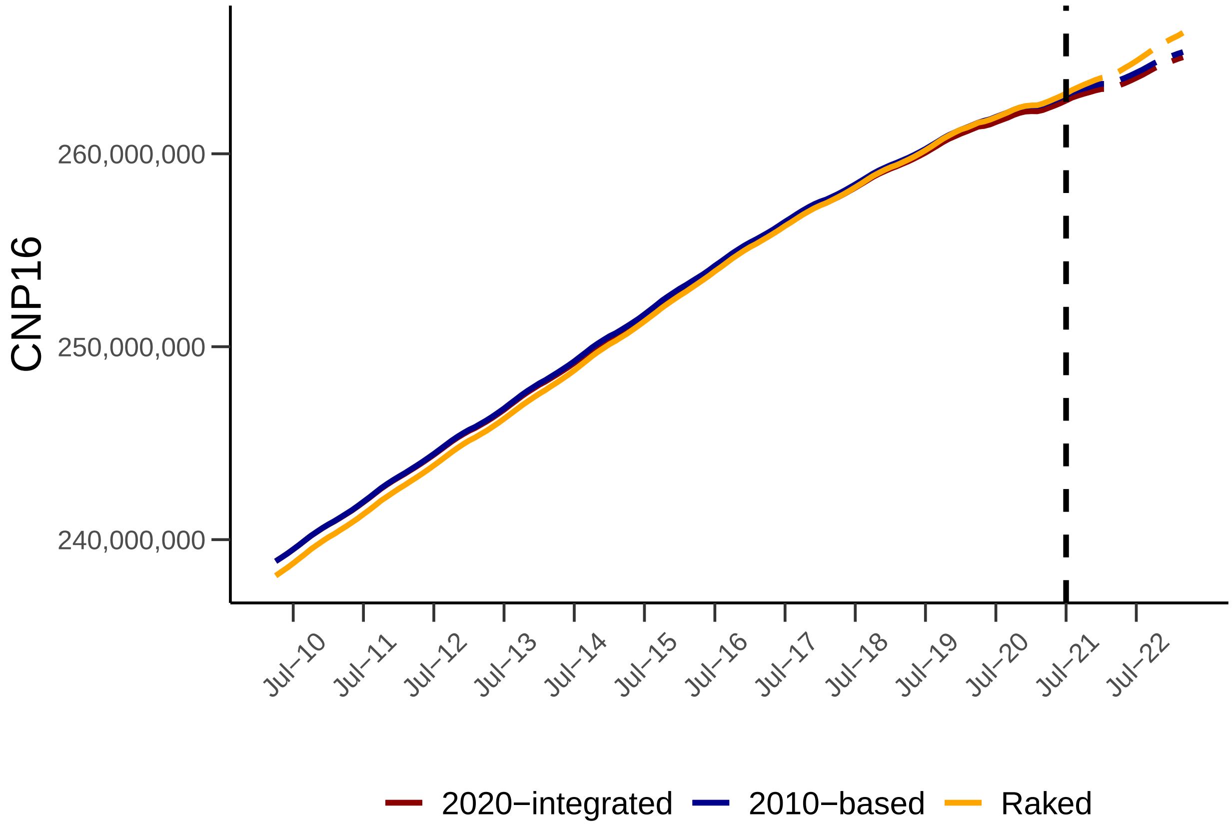

Figure A 4: Effect of Using 2010-Based Institutional Prevalence Rates v. 2010 and 2020 Rates

Notes: Figure shows the civilian noninstitutional population estimated using institutional-prevalence rates derived from the 2010 census alone and interpolating the 2010 and 2020 rates. Data from the 2010 census are from Summary File 1 and data from the 2020 census are from the Demographic and Housing Characteristics (DHC) file.

D Additional Tables

Table A 1: Comparing CNP16 with the ACS

Year	Counties	ACS (1yr)	CCM (raw)	CCM (adj)	MAPE (raw)	MAPE (adj)
2010	89	93,167,959	93,380,301	93,095,865	0.23%	0.08%
2011	93	97,197,163	97,509,156	97,252,892	0.32%	0.06%
2012	93	98,233,105	98,579,072	98,356,332	0.35%	0.13%
2013	99	102,968,137	103,444,327	103,238,132	0.46%	0.26%
2014	100	104,933,562	105,215,158	105,040,723	0.27%	0.10%
2015	102	107,256,213	107,439,775	107,323,367	0.17%	0.06%
2016	103	108,310,428	108,972,663	108,898,032	0.61%	0.54%
2017	104	110,266,617	110,533,402	110,489,834	0.24%	0.20%
2018	106	111,643,863	112,183,781	112,187,494	0.48%	0.48%
2019	107	112,659,748	113,522,413	113,575,043	0.76%	0.81%
2021	108	114,730,408	114,937,952	115,091,195	0.18%	0.31%

Notes: Table shows the differences in aggregate CNP16 from the ACS 1-year Table S1811 and the raw and adjusted cohort component estimates as of July 1st each year. The second column shows the number of published counties matched in the ACS 1-year data. Adjusted data are raked to the statewide CNP16 data from the Current Population Survey (CPS). 2020 ACS data for counties were not published due to data collection issues resulting from the COVID-19 pandemic.

E References

Beers, H. S. (1945). Six-term formulas for routine actuarial interpolation. *The Record of the American Institute of Actuaries*, *33*(2), 245–260.

Bryan, T. (2004). Population estimates. In J. S. Siegel & D. A. Swanson (Eds.), *The Materials and Methods of Demography* (2nd ed.). Elsevier.

U.S. Census Bureau. (2021). *Population controls for the 2020 ACS*. U.S. Census Bureau; 2010 Census of Population and Housing. https://www.census.gov/programs-surveys/acs/technical-documentation/user-notes/2021-01.html